



UNIVERSITÀ  
DEGLI STUDI  
DI UDINE

## Università degli studi di Udine

Machine Sounds, Sound Machines. Proceedings of the 22nd Colloquium on Musical Informatics

*Original*

*Availability:*

This version is available <http://hdl.handle.net/11390/1148323> since 2019-10-08T16:42:48Z

*Publisher:*

DADI - Dip. Arti e Design Industriale - Università IUAV

*Published*

DOI:

*Terms of use:*

The institutional repository of the University of Udine (<http://air.uniud.it>) is provided by ARIC services. The aim is to enable open access to all the world.

*Publisher copyright*

(Article begins on next page)

# Machine Sounds, Sound Machines

XXII Colloquio di Informatica Musicale XXII Colloquium of Music Informatics

Udine 20-23 Novembre 2018  
*Udine, November 20-23, 2018*



XXII Colloquio di Informatica Musicale  
*22nd Colloquium on Music Informatics*

Atti della Conferenza  
*Conference Proceedings*

---

Udine 20-23 Novembre 2018  
*Udine, November 20-23, 2018*

AA.VV.

MACHINE SOUNDS, SOUND MACHINES  
XXII Colloquio di Informatica Musicale  
*22nd Colloquium on Music Informatics*

**A cura di:**

**Editors:**

Federico Fontana, Andrea Gulli

© 2018 AIMI - Associazione Informatica Musicale Italiana

I diritti dei singoli articoli rimangono agli autori

*The copyrights of the single articles remain with the authors*

**ISBN 9788890341359**

**ISSN 2611-7355**

Per gentile collaborazione di Davide Rocchesso e Stefano delle Monache

*Courtesy of Davide Rocchesso and Stefano delle Monache*

**Publicato da:**

**Publisher:**

DADI - Dip. Arti e Design Industriale

Università IUAV di Venezia

**In co-edizione con:**

**Co-published with:**

Università degli Studi di Udine

**Riferimento BibTeX:**

```
@proceedings{XXIICIM,  
Editor = {Federico Fontana and Andrea Gulli},  
Organization = {AIMI - Associazione Informatica Musicale Italiana},  
Publisher = {DADI - Dip. Arti e Design Industriale. Università IUAV  
di Venezia},  
Title = {Machine Sounds, Sound Machines. Atti del XXII CIM - Colloquio  
di Informatica Musicale},  
Year = {2019}  
}
```



Quest'opera è sotto licenza Creative Commons Attribuzione - Non opere derivate 4.0 Unported License.

Copia della licenza è disponibile presso <http://creativecommons.org/licenses/by-nd/4.0/>.

*This work is licensed under the Creative Commons Attribution-NoDerivs 4.0 Unported License.*

*To view a copy of this license, visit <http://creativecommons.org/licenses/by-nd/4.0/>.*

*Jean-Claude Risset  
(1938 - 2016)*



*(Photo: Laura Zattra)*



ASSOCIAZIONE INFORMATICA MUSICALE ITALIANA



**UNIVERSITÀ  
DEGLI STUDI  
DI UDINE**

**hic sunt futura**

**CENTRO  
POLIFUNZIONALE  
DI PORDENONE**



**DIPARTIMENTO DI  
SCIENZE MATEMATICHE  
INFORMATICHE E FISICHE**



**CONSERVATORIO  
STATALE DI MUSICA  
JACOPO TOMADINI  
UDINE**

COMUNE DI

**udine**



**FONDAZIONE  
FRIULI**



**REGIONE AUTONOMA  
FRIULI VENEZIA GIULIA**

XXII COLLOQUIO DI INFORMATICA MUSICALE  
22nd COLLOQUIUM ON MUSIC INFORMATICS

L'appuntamento del CIM rappresenta da oltre quarant'anni un'imprescindibile occasione di confronto per la comunità dell'informatica musicale italiana. Il tema di quest'anno ci invita a riflettere sulla reciprocità della relazione suono/macchina e sulle molteplici implicazioni che hanno riscontro nella nostra esperienza quotidiana. Verranno presentate 23 composizioni elettroacustiche tra acustiche, miste e videomusicali, 5 installazioni e ancora 4 workshop, 3 keynote speech, 44 comunicazioni scientifiche articolate in 3 sessioni demo, 3 sessioni poster e 4 sessioni speciali. Una novità di questo CIM è l'introduzione del premio Teresa Rampazzi alla più originale composizione elettroacustica selezionata dalla call for music, riconoscimento che affiancherà il premio Aldo Piccialli al contributo scientifico maggiormente innovativo nella ricerca sull'informatica musicale. Si è voluto così ricordare la figura della musicista veneta che con grande passione ha dato un contributo determinante alla formazione della prima generazione della musica informatica italiana. Il concerto inaugurale proporrà le musiche di Jean-Claude Risset a due anni dalla sua scomparsa, un sentito omaggio alla fondamentale figura del musicista-scienziato francese.

*For over forty years, the CIM has been an essential opportunity for the Italian music information technology community to meet and exchange ideas. This year's theme invites us to reflect on the reciprocity of the relationship between sound and machine and on the many implications that are reflected in our daily experience. 23 electroacoustic compositions will be presented, including acousmatic, mixed and video-musical, 5 installations and 4 more workshops, 3 keynote speeches, 44 scientific communications divided into 3 demo sessions, 3 poster sessions and 4 special sessions. A novelty of this CIM is the introduction of the Teresa Rampazzi prize to the most original electroacoustic composition selected from the call for music, an award that will accompany the Aldo Piccialli prize to the most innovative scientific contribution in the research on musical informatics. It was decided to remember the figure of the Venetian musician who, with great passion, made a decisive contribution to the formation of the first generation of Italian computer music. The opening concert will propose the music of Jean-Claude Risset two years after his death, a heartfelt tribute to the fundamental figure of the French musician-scientist.*

MACHINE SOUNDS, SOUND MACHINES

Il tema dell'edizione 2018 è il rapporto tra suono e macchina. Attraverso il suo inarrestabile sviluppo, la macchina infatti sta diventando protagonista in tutti i processi che coinvolgono il suono musicale. Fino a pochi anni fa limitati agli ambiti tradizionali della registrazione, sintesi, trattamento e riproduzione del suono, ultimamente questi processi sono cresciuti avvalendosi della raccolta, l'analisi e la diffusione capillare di grandi quantità di dati. La comunità dell'informatica musicale non è sorpresa da questo sviluppo a cui lei stessa ha contribuito, e tuttavia deve impegnarsi adesso per non smarrire il percorso artistico-creativo che fino a oggi ha garantito la positività ("soundness") delle macchine per il suono. Gli schemi adoperati in passato devono confrontarsi con un'intelligenza artificiale che pervasivamente assimila dati e poi, dopo averli trattati con algoritmi sempre più difficili da comprendere, contribuisce a dettare i gusti e le nuove frontiere della musica. Per questo, il XXII Colloquio di Informatica Musicale affiancherà la programmazione a tema le macchine per il suono con una serie di eventi didattici attorno allo stesso tema, preliminari al Colloquio. Questi eventi sono destinati ai giovani musicisti e in particolare agli studenti dei Conservatori, ai quali è consegnato il compito di comprendere più a fondo la macchina nella speranza che essi contribuiscano a tenere vivo il suo carattere "sound" come parte della loro attività artistico-creativa. Una speranza, questa, condivisa già nel 1991 da Jean-Claude Risset in un suo intervento sul Computer Music Journal, in cui auspicava un nuovo ruolo delle macchine per condividere la musica e suggerire soluzioni al musicista senza deprimerne la creatività. In quest'ottica di condivisione, il XXII Colloquio di Informatica Musicale ricorderà il compositore a un anno dalla scomparsa nel concerto inaugurale, accogliendo inoltre contributi tematici a lui dedicati.

*The general theme of the 2018 edition is the relationship between sound and machine. Through their ever-growing development, machines are becoming protagonists in every process involving musical sound. Limited until few years ago to the traditional sound recording, synthesis, processing and reproduction stages, recently their role has tremendously expanded thanks to the possibility to collect, analyse and widely spread large amounts of data. The computer science community has not been surprised by this growth, indeed fueled by the own activity of its researchers, and yet this community must act now not to lose the artistic and creative roadmap that so far has kept such machines soundness stable. The old approaches must face an artificial intelligence that acquires data pervasively and then, once such data have been processed through algorithms which are harder and harder to understand, prominently dictates the taste and the new frontiers of music. For this reason the XXII Colloquium of Musical Informatics will enrich its program about sound machines with a list of preliminary tutorials around the same argument. Such events are targeted to an audience counting young musicians, and especially students from the Schools of music, whose artistic and creative activity is hoped to be part of a deeper understanding of the machine aimed at keeping its sound behavior. This hope was shared as early in 1991 by Jean-Claude Risset who, in a Computer Music Journal contribution, fostered a new role of the machine concerning music sharing and for suggesting solutions to musicians without depressing their creativity. With the same spirit of sharing, the XXII Colloquium of Musical Informatics will tribute part of the opening concert to the composer one year after his death, and furthermore welcome thematic works in his memory.*

**Coordinamento generale:**

Federico Fontana

**Coordinamento scientifico:**

Federico Avanzini

Stefano Delle Monache

Carlo Drioli

**Coordinamento programma musicale:**

Stefano Bassanese

Luca Cossettini

Roberto Girolin

**Coordinamento musicale, artistico e organizzativo locale:**

Roberto Barbieri

Virginio Zoccatelli

**Comitato di programma:**

Federico Avanzini, Università di Milano

Roberto Barbieri, Conservatorio di Udine

Stefano Bassanese, Conservatorio di Torino

Davide Bonsi, Electrolux SpA

Luca Cossettini, Università di Udine

Stefano Delle Monache, Università IUAV di Venezia

Carlo Drioli, Università di Udine

Federico Fontana, Università di Udine

Roberto Girolin, Pordenone

Giorgio Klauer, Conservatorio di Venezia

Renato Miani, Conservatorio di Udine

Mario Pagotto, Conservatorio di Udine

Anna Terzaroli, Conservatorio di Roma

Andrea Valle, Università di Torino

Paolo Zavagna, Conservatorio di Venezia

Virginio Zoccatelli, Conservatorio di Udine

**Comitato di revisione (programma scientifico):**

Giacomo Albert

Federico Altieri

Federico Avanzini

Adriano Barà

Alberto Bernardini

Michele Buccoli

Stefano d'Angelo

Amalia de Götzen

Giovanni de Poli

Fabio de Sanctis de Benedictis

Alan del Piccolo

Michele del Prete

Stefano delle Monache

Carlo Drioli

Michele Ducceschi

Federico Fontana

Leonardo Gabrielli

Michele Geronazzo

Bruno Giordano

Vincenzo Lombardo

Luca Ludovico

Enrico Marchetto

Davide Andrea Mauro

Fabio Morreale

Nicola Orio

Stefano Papetti

Sandra Pauletto

Niccol Pretto

Davide Rocchesso

Antonio Rodà

Stefania Serafin

Simone Spagnol

Andrea Valle

Domenico Vicinanza

Gualtiero Volpe

Stefano Zambon

Laura Zattra

**Comitato di revisione (programma musicale):**

Andrea Agostini

Stefano Bassanese

Tonino Battista

Laura Bianchini

Nicola Buso

Lelio Camilleri

Luigi Ceccarelli

Antonino Chiaramonte

Walter Ciancusi

Fabio Cifariello Ciardi

Luca Cossettini

Franco Degrossi

Daniele Ghisi

Roberto Girolin

Giorgio Klauer

Carlo Laurenzi

Eric Maestri

Massimo Marchi

Alberto Novello

Alessandro Olto

Paolo Pachini

Pietro Polotti

Riccardo Santoboni

Anna Terzaroli

Massimiliano Viel

Roberto Zanata

Paolo Zavagna

**Coordinamento installazioni:**

Yuri De Pra

**Regia in Sala concerti:**

Nicola Buso

**Assistenza tecnica concerti:**

Samuel Montalto

Mattia Piani

**Realizzazioni grafiche e layout:**

Alessandro Olto

**Collaborazione alle realizzazioni grafiche:**

Andrea Boato

Mettheew Chivilò

Nicolas Duri

Alessandra Marchetto

Davide Pasqual

Giovanni Scaramucci

|   |     |
|---|-----|
| <b>Programma del colloquio</b><br><b>Colloquium program (in Italian)</b>  | 7   |
| <b>Indice degli abstract</b><br><b>Index of abstracts</b>   | 13  |
| <b>Indice degli articoli</b><br><b>Index of papers</b>  |     |
| STEFANO PAPETTI<br><i>Aptica e Interazione Musicale</i>   | 27  |
| FEDERICA BRESSAN<br><i>Dynamic Preservation of Interactive Art: The next frontier of Multimedia Cultural Heritage (DaphNet)</i>   | 35  |
| FABIO MORREALE<br><i>A Holistic Approach to the Design of Digital Musical Instruments</i>   | 37  |
| MICHELE DUCCESCHI<br><i>Un Modello Fisico per Pianoforte Preparato</i>  | 39  |
| SIMONE SPAGNOL, KRISTJÁN BJARKI PURKHÚS, SVERRIR KARL BJÖRNSSON AND RÚNAR UNNTHÓRSSON<br><i>Misure di HRTF su una testa KEMAR con padiglioni auricolari intercambiabili</i> | 47  |
| MICHELE GERONAZZO<br><i>Tecnologie per l'Interazione Sonora in Contesti di Realtà Virtuale e Aumentata Immersiva</i>  | 53  |
| STEFANO D'ANGELO<br><i>Lightweight Virtual Analog Modeling</i>  | 59  |
| LEONARDO GABRIELLI, FABIO VESPERINI, DIEGO DROGHINI AND STEFANO SQUARTINI<br><i>Rima Glottidis: Experimenting Generative Raw Audio Synthesis for a Sound Installation</i>   | 64  |
| LUCA DANIELI<br><i>Points of discontinuity and spatial cadences: two experimental concepts for the development of musical theory</i>  | 70  |
| ADRIANO CASTALDINI<br><i>Vexations: La Serie. Dati dal corpo, interpretazione musicale e spazio d'ascolto.</i>  | 77  |
| LUCA MORINO AND DANIELE GHISI<br><i>Applicazioni storiche della Teoria del Caos nella computer music</i>  | 85  |
| STEFANO DELLE MONACHE AND DAVIDE ROCCHESSE<br><i>To embody or not to embody: A sound design dilemma</i>   | 93  |
| DAVIDE ROCCHESSE AND MARIA MANNONE  | 101 |

*Embryo of a Quantum Vocal Theory of Sound*

|  |     |
|--|-----|
| ANTONIO MASTOGIACOMO<br><i>Sulla musica meccanica a partire da John Philip Sousa</i>   | 109 |
| FEDERICO AVANZINI, ADRIANO BARATÈ, GOFFREDO HAUS, LUCA ANDREA LUDOVICO, DAVIDE ANDREA MAURO, STAVROS NTALAMPIRAS AND GIORGIO PRESTI<br><i>Quale futuro per il formato IEEE 1599</i>  | 115 |
| FABIO DE SANCTIS DE BENEDICTIS<br><i>FDSDB XXth CT: una libreria di composizione algoritmica per Open Music e PWGL dedicata alle tecniche compositive del Novecento</i>  | 122 |
| ANDREA VALLE<br><i>SamPComp: sample-based techniques for algorithmic composition</i>   | 128 |
| GIORGIO KLAUER<br><i>Il prototipo "scuba": un indumento aumentato per progettare l'interazione sonora con i vestiti</i>  | 136 |
| MICHELE GERONAZZO, STEFANIA SERAFIN, ROLF NORDAHL, EDOARDO DEGLI INNOCENTI, DIEGO VESCOVI AND FEDERICO AVANZINI<br><i>Tecnologie per la didattica musicale: un'esperienza con la realtà virtuale</i>                                     | 144 |
| STEFANO DELLE MONACHE, MICHELE BUCCOLI, LUCA COMANDUCCI, AUGUSTO SARTI, GIOVANNI COSPITO, ENRICO PIETROCOLA AND FILIPPO BERBENNI<br><i>Time is not on my side: network latency, presence and performance in remote music interaction</i> | 152 |
| MARTINO SAROLLI AND ANDREA AGOSTINI<br><i>Deep Learning per l'Elaborazione Digitale del Suono: possibilità di integrazione nella prassi compositiva elettroacustica</i>  | 160 |
| LAURA ZATTRA, NICOLAS MISDARIIS, FRANK PECQUET, NICOLAS DONIN AND DAVID FIERRO<br><i>Analysis of Sound Design Practicies. Research Methodology</i>   | 168 |
| SANDRA PAULETTO AND BART WALUS<br><i>Reflecting on the Role of Sound in an Immersive Multi-media Intervention for Health Communication</i>   | 176 |

**Indice dei poster**

**Index of posters**

|  |     |
|--|-----|
| GIOVANNI ALBINI<br><i>bach: A Procedural Video Game Featuring Music Visualization for an Interdisciplinary Computer Based Approach to the Mathemusical Education</i>       | 187 |
| ALBERTO NOVELLO, ALESSANDRO FIORELMONDO AND ARCANGELO DIDONATO<br><i>PhOS: light-sensitive gestural surface</i>  | 189 |
| ARCANGELO DI DONATO AND GIORGIO KLAUER<br><i>Il suono attuale dei luoghi perduti</i>   | 191 |
| ARCANGELO DI DONATO, ALBERTO NOVELLO AND ALESSANDRO FIORELMONDO<br><i>Il controller pittorico: uno strumento gestuale per la sintesi sonora</i>                            | 197 |
| ALESSANDRO BARILE, SAMUEL MONTALTO AND LUCA CIGAINA<br><i>Safe and Sound</i>   | 202 |
| LORENA CANTARUT<br><i>LIS-10 e Heed: due applicazioni di Shape Memory Alloys nell'ambito della Sound Art</i>   | 205 |
| MICHELA COPPOLA<br><i>MYTRENDS: playing words</i>  | 209 |
| KRISTJAN STOPAR, ROBERTO CALABRETTO AND LUCA COSSETTINI<br><i>Le nuove scritture musicali per il cinema. Studi di registrazione, media digitali e pratiche compositive</i> | 215 |
| LUCA RICHELLI<br><i>Motion Capture e Strumenti Aumentati: Controllo Gestuale e Gesto Compositivo</i>   | 220 |
| YURI DE PRA, FEDERICO FONTANA AND MICHELE SIMONATO<br><i>Development of real-time audio applications using Python</i>  | 226 |

|   |     |
|---|-----|
| ARCANGELO DI DONATO, ALBERTO NOVELLO AND GIORGIO KLAUER<br><i>Il gioiello elettroacustico</i>               | 232 |
| ANDREA BOLZONI AND STEFANO CATENA<br><i>Distorsione di massa</i>  | 237 |
| GIOVANNI SPARANO<br><i>GrainLab - Software open source per la sintesi granulare quasi-sincrona</i>          | 243 |
| PAOLO ZAVAGNA, FRANCESCO PECCOLO AND RICCARDO SELLAN<br><i>Electroacoustic and Computer Music Timelines</i> | 246 |
| GIACOMO LEPRI<br><i>InMuSIC: un sistema interattivo multimodale per l'improvvisazione elettroacustica</i>   | 249 |





Programma del colloquio  
*Colloquium program*



**MARTEDÌ 20 NOVEMBRE 2018**

TUESDAY, NOVEMBER 20, 2018

*Risposta all'impulso: l'informatica musicale spiegata agli studenti di musica non elettronica*

09:30 Sala Vivaldi  
**Workshop 1**

LUCA COSSETTINI  
*Musica ex machina*

11:00 Sala Vivaldi  
**Workshop 2**

ROBERTO GIROLINI  
*Scolpire il suono (natura, distorsione e trama narrativa): come orientarsi?*

13:00 Pausa pranzo

14:00 Sala Vivaldi  
**Workshop 3**

STEPHEN BARRASS  
*Make your own Sound Machine with MozziByte*

15:30 Sala Vivaldi  
**Workshop 4**

ANDREA VALLE  
*Introduzione alla composizione algoritmica con SuperCollider*

18:00 - Sala Udienze  
**Concerto off-CIM**

Brani in prima esecuzione assoluta degli allievi del Conservatorio di Udine

SILVIA VALENTINI  
*Strumento uomo (2018)*  
per clarinetto basso, clarinetto in Sib ed elettronica  
clarinetto basso: Eric Zanella  
clarinetto: Giacomo Cozzi  
regia del suono: Silvia Valentini  
LORENZO GIOCO  
*Klimen en Dalen (2018)*  
per due clarinetti ed elettronica  
clarinetti: Giacomo Cozzi, Eric Zanella  
regia del suono: Lorenzo Gioco

INGRID MACUS  
*Jezero (2018)*  
per due voci, pianoforte, percussione ed elettronica  
soprano: Nataša Trcek  
mezzosoprano: Mirjam Jelnikar  
pianoforte: Ingrid Macus  
percussione: Alberto Zenarolla  
regia del suono: Marco Sgobino

20:45 Teatro San Giorgio  
**Concerto inaugurale**

*Omaggio a Jean-Claude Risset*

JEAN CLAUDE RISSET  
*Mutations (1969)*  
versione per il film di Lillian Schwartz (1973)  
regia del suono: Paolo Zavagna  
JEAN CLAUDE RISSET  
*Variants (1994)*  
per violino e live electronics  
violino: Alessandro Fagioli  
live electronics: Paolo Zavagna  
JEAN CLAUDE RISSET  
*Resonant Sound Spaces (2001-2002)*  
1. *Bells Brass Metal*  
2. *Filters*  
3. *Plucked*  
4. *Reverberated*  
5. *Horn Bells*

acusmatico  
regia del suono: Paolo Zavagna  
JEAN CLAUDE RISSET  
*Dialogues (1975)*  
per quattro strumentisti e nastro magnetico  
flauto: Giorgio Marcossi  
clarinetto: Nicola Bulfone  
percussione: Roberto Barbi-eri  
pianoforte: Paolo Chiarandini  
regia del suono: Stefano Basanese

**MERCOLEDÌ 21 NOVEMBRE 2018**

WEDNESDAY, NOVEMBER 21, 2018

09:30 Sala Vivaldi  
**Saluti istituzionali**

9:50 Sala Vivaldi  
**Sessione 1**

*Representing sounds, sound representations*

Chair  
STEFANO DELLE MONACHE

LUCA DANIELI  
*Points of discontinuity and spatial cadences: two experimental concepts for the development of musical theory*

ADRIANO CASTALDINI  
*Vexations: La Serie. Dati dal corpo, interpretazione musicale e spazio d'ascolto*

LUCA MORINO AND DANIELE GHISI  
*Applicazioni storiche della Teoria del Caos nella computer music*  
STEFANO DELLE MONACHE AND DAVIDE ROCCHESSE  
*To embody or not to embody: A sound design dilemma*  
DAVIDE ROCCHESSE AND MARIA MANNONE  
*Embryo of a Quantum Vocal Theory of Sound*

11:30 Pausa caffè

12:00 Sala Vivaldi  
**Keynote 1**

Chair  
DAVIDE ROCCHESSE

STEPHEN BARRASS  
*Sonic Information Design*

12:50 Sala Vivaldi  
**Poster craze 1**

13:00 Pausa pranzo

14:00 Ingresso via Treppo  
**Sessione poster/demo 1**

Chair  
SIMONE SPAGNOL

GIOVANNI ALBINI  
*bach: A Procedural Video Game Featuring Music Visualization for an Interdisciplinary Computer Based Approach to the Mathematical Education*

ALBERTO NOVELLO, ALESSANDRO FIORDELMONDO AND ARCANGELO DIDONATO  
*PhOS: light-sensitive gestural surface*

ARCANGELO DI DONATO AND GIORGIO KLAUER  
*Il suono attuale dei luoghi perduti*

ARCANGELO DI DONATO, ALBERTO NOVELLO AND ALESSANDRO FIORDELMONDO  
*Il controller pittorico: uno strumento gestuale per la sintesi sonora*

ALESSANDRO BARILE, SAMUEL MONTALTO AND LUCA CIGAINA  
*Safe and Sound*

15:00 Sala Vivaldi  
**Sessione speciale**

*Informatici musicali all'estero 1*

Chair  
STEFANO D'ANGELO

STEFANO PAPETTI  
*Aptica e Interazione Musicale*

FEDERICA BRESSAN  
*Dynamic Preservation of Interactive Art: The next frontier of Multimedia Cultural Heritage (DaphNet)*

FABIO MORREALE  
*A Holistic Approach to the Design of Digital Musical Instruments*  
MICHELE DUCCESCHI  
*Un Modello Fisico per Pianoforte Preparato*

**MERCOLEDÌ 21 NOVEMBRE 2018**

WEDNESDAY, NOVEMBER 21, 2018

16:20 Sala Vivaldi  
**Sessione d'ascolto 1**

MAURIZIO DI BERARDINO  
*The disgraceful displacement of a personality shift (2015)*  
acusmatico  
NICOLA FUMO FRATTEGANI  
*Banlieue cuivrée (2018)*  
acusmatico

16:40 Sala Vivaldi  
**Sessione speciale**

*Informatici musicali all'estero 2*

*Chair*  
FABIO MORREALE

SIMONE SPAGNOL, KRISTJÁN BJARKI PURKHÚS, SVERRIR KARLBJÖRNSSON AND RÚNAR UNNTHÓRSSON  
*Misure di HRTF su una testa KEMAR con padiglioni auricolari intercambiabili*  
MICHELE GERONAZZO  
*Tecnologie per l'Interazione Sonora in Contesti di Realtà Virtuale e Aumentata Immersiva*  
STEFANO D'ANGELO  
*Lightweight Virtual Analog Modeling*  
LEONARDO GABRIELLI, FABIO VESPERINI, DIEGO DROGHINI AND STEFANO SQUARTINI  
*Rima Glottidis: Experimenting Generative Raw Audio Synthesis for a Sound Installation*

20:45 Teatro San Giorgio  
**Concerto CIM n. 1**

ROBERTO ZANATA  
*Meccanica I (2017)*  
per flauto basso e live electronics  
flauto basso: Giorgio Marcossi  
MARCO MARINONI  
*Siamo sporchi di guerra (2017-2018)*  
acusmatico  
FABIO DE SANCTIS DE BENEDICTIS  
*Balaeonoptera (2015)*  
per clarinetto basso e live electronics  
clarinetto basso: Nicola Bulfone  
MARTINO SAROLLI  
*Parafulmini e parassiti (2018)*  
acusmatico  
NICOLETTA ANDREUCCETTI, LORENZO BALLERINI  
*Resounding Resonances (2017)*  
per flauto (piccolo), clarinetto in Sib, pianoforte e live electronics  
flauto piccolo: Giorgio Marcossi  
clarinetto: Nicola Bulfone  
pianoforte: Paolo Chiarandini  
live electronics: Lorenzo Ballerini  
FEDERICO TESTONI  
*Kairos e Airon (2018)*  
acusmatico  
ALESSANDRO FIORDELMONDO, ARCANGELO DI DONATO  
*Il controller pittorico (2017)*  
performers: Alessandro Fiordelmondo e Arcangelo Di Donato

Ove non specificato, regia del suono e live electronics a cura degli autori

**GIOVEDÌ 22 NOVEMBRE 2018**

THURSDAY, NOVEMBER 22, 2018

09:30 Sala Vivaldi  
**Sessione 2**

*Interactive sounds, sound interactions*

*Chair*  
FEDERICA BRESSAN

ANTONIO MASTOGIACOMO  
*Sulla musica meccanica a partire da John Philip Sousa*  
FEDERICO AVANZINI, ADRIANO BARATÈ, GOFFREDO HAUS, LUCA ANDREA LUDOVICO, DAVIDE ANDREA MAURO, STAVROS NTALAMPIRAS AND GIORGIO PRESTI  
*Quale futuro per il formato IEEE 1599*  
FABIO DE SANCTIS DE BENEDICTIS  
*FDSDB XXth CT: una libreria di composizione algoritmica per Open Music e PWGL dedicata alle tecniche compositive del Novecento*  
ANDREA VALLE  
*SampComp: sample-based techniques for algorithmic composition*  
GIORGIO KLAUER  
*Il prototipo "scuba": un indumento aumentato per progettare l'interazione sonora con i vestiti*

11:10 Sala Vivaldi  
**Sessione d'ascolto 2**

STEFANO GIACOMELLI  
*l.Side #2 (Flow of Sensations) (2018)*  
acusmatico  
COSTANTINO RIZZUTI  
*Due studi su Le Corde dell'Anima (2018)*  
acusmatico

11:30 Pausa caffè

12:00 Sala Vivaldi  
**Keynote 2**

*Chair*  
FEDERICO AVANZINI

STEFANIA SERAFIN  
*Sonic interactions for all*

12:50 Sala Vivaldi  
**Poster craze 2**

13:00 Pausa pranzo

14:00 Ingresso via Treppo  
**Sessione poster/demo 2**

*Chair*  
ALESSANDRO OLTO

LORENA CANTARUT  
*LIS-10 e Heed: due applicazioni di Shape Memory Alloys nell'ambito della Sound Art*  
MICHELA COPPOLA  
*MYTRENDS: playing words*  
KRISTJAN STOPAR, ROBERTO CALABRETTO AND LUCA COSSETTINI  
*Le nuove scritture musicali per il cinema. Studi di registrazione, media digitali e pratiche compositive*  
LUCA RICHELLI  
*Motion Capture e Strumenti Aumentati: Controllo Gestuale e Gesto Compositivo*  
YURI DE PRA, FEDERICO FONTANA AND MICHELE SIMONATO  
*Development of real-time audio applications using Python*

15:00 Sala Vivaldi  
**Sessione speciale**

*Distinguished coders 1*

*Chair*  
STEFANIA SERAFIN

ROMAIN MICHON  
*A quick Faust tutorial*

**GIOVEDÌ 22 NOVEMBRE 2018**  
THURSDAY, NOVEMBER 22, 2018

**VENERDÌ 23 NOVEMBRE**  
FRIDAY, NOVEMBER 23, 2018

|   |   |   |   |
|---|---|---|---|
| <p>16:20 Sala Vivaldi<br/><b>Sessione d'ascolto 3</b></p> <p>MASSIMO VITO AVANTAGGIATO<br/><i>Atlas of uncertainty (2016)</i><br/>opera audiovisiva<br/>ANNA TERZAROLI<br/><i>Dark Path #4 (2015)</i><br/>acusmatico</p> <p>16:40 Sala Vivaldi<br/><b>Assemblea AIMI</b></p> <p>A seguire Sala Vivaldi<br/><b>Direttivo AIMI</b></p> <p>20:45 Teatro San Giorgio<br/><b>Concerto CIM n. 2</b></p> <p>NICOLA BARONI<br/><i>K_messages (2014-2016)</i><br/>1. <i>Vor dem Gesetz</i><br/>2. <i>The Wish to be a Red Indian</i><br/>per ipercello (violoncello e live electronics)<br/>ipercello: Nicola Baroni<br/>ANTONIO D'AMATO<br/><i>Béatitudes lysergiques (2018)</i><br/>acusmatico<br/>DAVIDE WANG<br/><i>Un pensiero... (2018)</i><br/>per clarinetto in Sib e live electronics<br/>clarinetto: Davide Teodoro<br/>GIOVANNI SPARANO<br/><i>Cosmos (2018)</i><br/>acusmatico<br/>LUCA RICHELLI<br/><i>Gesti di tensione (2017)</i><br/>per violino ed elettronica<br/>violino: Andrea Scaramella<br/>LUCA MORINO<br/><i>My Taste Passes Through My Stomach (2018)</i><br/>acusmatico<br/>ROBERTO DOATI E PAOLO PACHINI<br/><i>Seppie senz'osso (2013)</i><br/>opera video</p> <p>Ove non specificato, regia del suono e live electronics a cura degli autori</p> | <p>09:30 Sala Vivaldi<br/><b>Sessione 3</b></p> <p><i>Learning sounds, sound learning</i></p> <p><i>Chair</i><br/>SANDRA PAULETTO</p> <p>MICHELE GERONAZZO, STEFANIA SERAFIN, ROLF NORDAHL, EDOARDO DEGLI INNOCENTI, DIEGO VESCOVI AND FEDERICO AVANZINI<br/><i>Tecnologie per la didattica musicale: un'esperienza con la realtà virtuale</i><br/>STEFANO DELLE MONACHE, MICHELE BUCCOLI, LUCA COMANDUCCI, AUGUSTO SARTI, GIOVANNI COSPITO, ENRICO PIETROCOLA AND FILIPPO BERBENNI<br/><i>Time is not on my side: network latency, presence and performance in remote music interaction</i><br/>MARTINO SAROLLI AND ANDREA AGOSTINI<br/><i>Deep Learning per l'Elaborazione Digitale del Suono: possibilità di integrazione nella prassi compositiva elettroacustica</i><br/>LAURA ZATTRA, NICOLAS MISDARIIS, FRANK PECQUET, NICOLAS DONIN AND DAVID FIERRO<br/><i>Analysis of Sound Design Practicies. Research Methodology</i><br/>SANDRA PAULETTO AND BART WALUS<br/><i>Reflecting on the Role of Sound in an Immersive Multi-media Intervention for Health Communication</i></p> | <p>11:10 Sala Vivaldi<br/><b>Sessione d'ascolto 4</b></p> <p>ROCÍO CANO VALIÑO<br/><i>Astérian (2018)</i><br/>acusmatico<br/>ANTONIO SCARCIA<br/><i>Interludio: "And I rejoice in my bonie Bell" (2018)</i><br/>acusmatico</p> <p>11:30 Pausa caffè</p> <p>12:00 Sala Vivaldi<br/><b>Keynote 3</b></p> <p><i>Chair</i><br/>STEFANO BASSANESE</p> <p>CARMINE EMANUELE CELLA<br/><i>Can Picasso think in shapes?</i></p> <p>12:50 Sala Vivaldi<br/><b>Poster craze 3</b></p> <p>13:00 Pausa pranzo</p> <p>14:00 Ingresso via Treppo<br/><b>Sessione poster/demo 3</b></p> <p><i>Chair</i><br/>MICHELE GERONAZZO</p> <p>ARCANGELO DI DONATO, ALBERTO NOVELLO AND GIORGIO KLAUER<br/><i>Il gioiello elettroacustico</i><br/>ANDREA BOLZONI AND STEFANO CATENA<br/><i>Distorsione di massa</i><br/>GIOVANNI SPARANO<br/><i>GrainLab - Software open source per la sintesi granulare quasi-sincrona</i><br/>PAOLO ZAVAGNA, FRANCESCO PECCOLO AND RICCARDO SELLAN<br/><i>Electroacoustic and Computer Music Timelines</i><br/>GIACOMO LEPRI<br/><i>InMusIC: un sistema interattivo multimodale per l'improvvisazione elettroacustica</i></p> | <p>15:00 Sala Vivaldi<br/><b>Sessione speciale</b></p> <p><i>Distinguished coders 2</i></p> <p><i>Chair</i><br/>PIETRO POLOTTI</p> <p>ANDREA VALLE<br/><i>First SuperCollision with SuperCollider</i></p> <p>16:20 Sala Vivaldi<br/><b>Sessione d'ascolto 5</b></p> <p>TOMMASO MUTO<br/><i>For PIANO INTERO (2018)</i><br/>acusmatico</p> <p>16:40 Sala Vivaldi<br/><b>Premio Piccialli - Rampazzi</b></p> <p>A seguire Sala Vivaldi<br/><b>Riunione CoME</b></p> <p>17:30 Sala Udienze<br/><b>Concerto off-CIM</b></p> <p><i>La mimesi della natura nel suono strumentale</i></p> <p>Branzi per strumenti e per strumenti ed elettronica degli allievi dei Conservatori di Udine e Trento-Riva del Garda</p> <p>ANDREA LORUSSO*<br/><i>Il volo dell'aquila (2018)</i><br/>per sassofono contralto, viola e pianoforte<br/>CLAUDIO DINA*<br/><i>38N 145W (2018)</i><br/>per sassofono contralto, viola e pianoforte<br/>ANDREA ALZETTA**<br/><i>The wrathful skies (2018)</i><br/>per sassofono soprano, viola, pianoforte ed elettronica</p> |
|---|---|---|---|

**VENERDÌ 23 NOVEMBRE**

FRIDAY, NOVEMBER 23, 2018

MARCO SGOBINO\*\*

*Viaggio verso il mare (2018)*

per sassofono contralto ed elettronica

MARCO FAGLIA\*

*Il risveglio del grillo (2018)*

per sassofono contralto, viola e pianoforte

INGRID MACUS\*\*

*Atollo (2018)*

per sassofono contralto, viola, pianoforte ed elettronica

MARCO RUSSO\*

*Animalis persona (2018)*

acusmatico

MICHELE CAGOL\*

*A whale called Nam Jr. (2018)*

per sassofono tenore, viola e pianoforte

FELICE DI PAOLO\*\*

*Long noises over a dismantled ostinato (2018)*

per sassofono baritono, viola, pianoforte ed elettronica

STEFANO WEGHER\*

*El grillo flipante (2018)*

per sassofono baritono

GIACOMO MAZZUCCHI\*

*Barbagianni (2018)*

per sassofono soprano e pianoforte

FABIO CONTI\*

*Il canto della notte (2018)*

per sassofono soprano, viola, pianoforte

sassofoni: Emanuele Dalmaso

viola: Andrea Mattevi

pianoforte: Marco Longo

regia del suono a cura degli autori

\* Conservatorio di Trento-Riva del Garda

\*\* Conservatorio di Udine

**INSTALLAZIONI**

INSTALLATIONS

Ingresso via Treppo

LORENA CANTARUT

*Heed (2018)*

scultura sonora

ENRICO DORIGATTI

*Life (2018)*

audiovisivo

FRANCESCO ROBERTO DANI

*PySCat (2017)*

demo performativa

MATTEO MARTINO

*CME2 (2017)*

ecosistema per suoni e luci

CESARE SALDICCO

*Still life (2012)*

audiovisivo

I seguenti lavori sono stati valutati positivamente dal Comitato di revisione, ma per ragioni diverse non è stato possibile inserirli nel programma.

ANTONIO D'AMATO

*Opus III (2015)*

acusmatico

FABIO DE SANCTIS DE BENE-DICTIS

*Fabula (2017)*

per sassofono baritono, elettronica e live electronics

MARCO MARINONI

*Eleusi (2015)*

per violino e live electronics

DEMIAN RUDEL REY

*Kh mia I (2016)*

per contrabbasso e live electronics

FEDERICO TESTONI

*Impressioni (2018)*

opera video



Abstract  
*Abstracts*



### **STEFANO PAPETTI**

Institute for Computer Music and Sound Technology, Zürcher Hochschule der Künste  
stefano.papetti@zhdk.ch

#### *Haptic and Musical Interaction*

During performance with a traditional musical instrument, a complex haptic (i.e. tactile and proprioceptive) interchange is usually established with the musician. Such exchange supports both the actual execution (e.g. by facilitating temporal, dynamic or pitch accuracy) and its aesthetic and expressive outcomes. Conversely, the physical experience currently offered by consumer electronics is extremely basic when compared to the richness of human interaction with the surrounding environment. In particular, improving haptic interaction with electronic musical devices - such as music controllers and digital musical instruments - still requires input from research in technology, perception and human-machine interaction. This paper summarizes the recent activity of the author in such research scenario.

### **FEDERICA BRESSAN**

IPEM - Institute for Psychoacoustics and Electronic Music Dept. of Art history, Music and Theatre studies Ghent University, Belgium  
federica.bressan@ugent.be

#### *Dynamic Preservation of Interactive Art: The next frontier of Multimedia Cultural Heritage (DaphNet)*

This paper presents an ongoing research project focussed on the preservation of interactive installation art, funded under the programme H2020-MSCA-IF-2015. The two-years project aims to define a framework that describes how people interact with interactive installations: the framework is based on empirical work on artistic interaction and combines qualitative and quantitative data. The ultimate goal is to contribute to the preservation and valorisation of interactive art, facilitating its replication, distribution and scholarly analysis. A major problem with the preservation of interactive art is that established archival practices require that objects are made static for preservation, failing to capture the essence of an art form that is dynamic and transitory. The case studies in this project are interactive installations developed in close collaboration with artists and researchers based in Ghent, Belgium. Because they were developed in the context of a scientific research lab, they feature built-in sensors that make the installation a work of art as well as a 'measuring tool' (in particular with the employment of the motion capture system).

### **FABIO MORREALE**

Augmented Instruments Lab Centre for Digital Music, Queen Mary University of London  
f.morreale@qmul.ac.uk

#### *A Holistic Approach to the Design of Digital Musical Instruments*

The number of new musical instruments has been rapidly growing in the last few years. Yet, few have achieved lasting impact beyond the first few performances. A few attempts have been made to solve this issue; most of them addressing it from a design. I argue that framing this issue only from a design perspective might fail to address the complexity of the problem at hand. Theory and praxis of musical interface design would benefit from a holistic approach, combining knowledge and methods from different disciplines, among which HCI, Music Cognition, and Philosophy.

### **MICHELE DUCCESCHI**

Acoustics and Audio Group, University of Edinburgh Scotland, UK  
michele.ducceschi@ed.ac.uk

#### *A Physical Model for Prepared Piano*

This work presents a numerical scheme for the simulation of a prepared piano. This mosaic instrument is composed of a normal piano, to which preparation instruments can be applied at will. These elements (rattles, shock absorbers, nails), supposedly punctiform, come into contact with the string for brief moments, enriching its harmonic content. Given the extremely non-linear nature of the system, an implicit number scheme will be proposed, the stability of which will be inferred from the conservation of a discreet, non-negative energy, similar to the Hamiltonian of the system.

### **SIMONE SPAGNOL**

Dip. di Ingegneria dell'Informazione, Università degli Studi di Padova, Padova, Italia  
spagnols@dei.unipd.it

### **KRISTJÁN BJARKI PURKHÚS, SVERRIR KARL BJÖRNSSON, RÚNAR UNNTHÓRSSON**

School of Engineering and Natural Sciences, University of Iceland, Reykjavík, Iceland  
runson@hi.is

#### *HRTF Measurements on a KEMAR Dummy-Head with Interchangeable Earphones*

This paper describes the acquisition of a dataset of Head-Related Transfer Function (HRTF) measurements performed at the University of Iceland. Full-sphere HRTFs (1513 positions) of a KEMAR mannequin with 20 different artificial pinnae obtained from molds of lifelike human heads are included. Preliminary analyses of results indicate the robustness and reproducibility of the acoustical measurements as well as their fidelity to reference KEMAR HRTFs. The purpose of the present data collection is to provide accurate input data for the implementation of a methodology for determining individual HRTF spectral patterns from a 3D representation of the user's pinna and, in general, to encourage investigations on the relation between HRTFs and anthropometric data through machine learning techniques.

### **MICHELE GERONAZZO**

Department of Architecture, Design and Media Technologies, Aalborg University, Copenhagen  
mge@create.aau.dk

### *Technologies for Sound Interaction in Immersive Virtual and Augmented Reality Environments*

This article deals with motivations, aims and states of progress of the two-year research project “Acoustically-trained 3D audio models for virtual reality applications developed” developed at the Department of Architecture, Design & Media Technologies - Aalborg University Copenhagen, started in September 2017. The development of technologies for an accurate and realistic multimodal rendering is pivotal element in proposing innovative strategies for applications that take advantage of real-time interaction in real, virtual (VR) or augmented reality (AR). This research project contributes to the next one generation of 3D audio frameworks that will be able to extract user and environmental acoustic information through microphones embedded in the headphones allowing: the training of (i) parametric models for user and room acoustics, (ii) dynamic compensation and acoustic transparency of headphones and (iii) definition of new interactions between user, device and environment. The fusion of virtual / real audio requires an innovative platform for personalized virtual acoustic reality. Here, research directions at the Multisensory Experience Lab will be summarized.

#### **STEFANO D’ANGELO**

Independent Researcher, Via P. Voso 22/B, Agropoli 84043, Italy  
s@dangelo.audio

### *Lightweight Virtual Analog Modeling*

The purpose of this paper is to give theoretical and practical advice to DSP programmers willing to implement efficient real-time circuit simulation algorithms. It summarizes the fundamental concepts behind the most widespread circuit modeling techniques and gives the reader sufficient references to get started. Then it tackles on a certain number of topics that are often neglected in theoretical studies yet highly relevant in real-world applications. Hopefully, this paper might also serve as a reference and an invite for researchers to direct their efforts towards problems and solutions that have a concrete impact.

#### **LEONARDO GABRIELLI, FABIO VESPERINI, DIEGO DROGHINI, STEFANO SQUARTINI**

Università Politecnica delle Marche, Ancona, Italy

### *Rima Glottidis: Experimenting Generative Raw Audio Synthesis for a Sound Installation*

Biologically-inspired algorithms such as artificial neural networks have been used extensively by computer music researchers, for generative music and algorithmic composition. The recent introduction of raw audio Machine Learning (ML) techniques, however, represents a significant leap because they seem to be able to learn both high-level (event) and low-level (timbre) information at once. Employing such techniques for creative purposes is very challenging at this early stage since there is lack of method, experience, and their computational cost is very high. In this paper, we describe the technical and creative process behind *Rima Glottidis*, an installation based on material from the homonym musical work from the artist *Økapi*, processed using raw audio ML techniques for an installation premiered at Blooming festival, Italy. Technical issues are described and critical aspects are reported as a reference for future projects.

#### **LUCA DANIELI**

University of Birmingham, Dept. of Electroacoustic Music Birmingham (UK)  
lxd430@bham.ac.uk  
dnllcu@gmail.com

### *Points of Discontinuity and Spatial Cadences: Two Experimental Concepts for the Development of Musical Theory*

In this article I aim to suggest a possible experimental approach to music theory for the development of new research and musical styles that connect traditional theory with the new compositional paradigms that characterize electroacoustic and computer musics. I highlight an interpretation of musical cadence as a point of discontinuity and present some preliminary results of a perceptual study conducted at the University of Birmingham on the relationship between music and space, which derives from this approach. Following the idea of cadence as a point of discontinuity, it has been possible to form a hypothesis on the possible perception of a sensation of cadence in relation to sound trajectories moving within a circular electroacoustic space. The cognitive hypothesis supposes the impossibility for the listeners to conceive a sound “object” passing “through” them, being themselves the discontinuity point of the spatial system in which they are placed. The study has involved 41 participants among experts and non-experts of electroacoustic music composition. Various sound stimuli diffused by use of multiple trajectories were presented to the participants, who were asked to evaluate their cadence level. A preliminary result reveals a tendency among electroacoustic music composers to consider sound stimuli ending at the opposite ends of a circular space as cadential.

#### **ADRIANO CASTALDINI**

Conservatorio di Musica “Giuseppe Tartini” di Trieste, Dipartimento di Nuovi Linguaggi Percussioni e Tecnologie Musicali, Scuola MNT (Musica e Nuove Tecnologie)  
adrjork@gmail.com

### *Vexations: The Series. Body Data, Musical Interpretation and Listening Space.*

*Vexations: The Series* is a performance for piano and sensor system, conceived and realized by Adriano Castaldini. The performance took place in Trieste in September 2016, and is based on the execution of the entire *Vexations* by Erik Satie (a 24 hours long piano piece) electronically augmented by EEG, EMG, GSR and temperature sensors worn by the pianist during the performance. *Vexations: The Series* is then a music interpretation experiment, where the psychophysical state of the pianist influences the timbre of an amplified piano, thanks to a real time interactive digital platform (based on Max 7) which converts psychophysical data into control values for the piano sound, making the pianist’s vexations audible. This research investigates the artistic efficacy of the body’s data used as an augmentation of the music interpretation (not simply of the instrument) and its influence on the physical listening space and the audience enjoyment.

#### **LUCA MORINO**

Conservatorio “G. Verdi” di Torino

morinoluc@gmail.com

**DANIELE GHISI**

Conservatorio "N. Paganini" di Genova

danieleghisi@bachproject.net

*Historical Applications of Chaos Theory in Computer Music*

This article presents a historical overview of the theoretical implications and practical applications that Chaos Theory has determined in the thought and creative processes of some avant-garde composers of the late twentieth century, particularly in the field of *computer music*, exposing some fundamental concepts of *deterministic chaos* and proposing a selection of composite techniques based on chaotic algorithms.

**STEFANO DELLE MONACHE**

Iuav University of Venice, Dept. of Architecture and Arts

stefano.dellemonache@iuav.it

**DAVIDE ROCCHESO**

University of Palermo, Dept. of Mathematics and Computer Science

davide.rocchesso@unipa.it

*To Embody or Not To Embody: A Sound Design Dilemma*

Designing with sound is about constructing appropriate sound representations of a concept, from the early ideas to the final product. A survey research on embodied sound sketching is presented, and the problems of early representation in sound design are discussed by analysing the questionnaire results of three workshops on vocal sketching.

**DAVIDE ROCCHESO**

University of Palermo

davide.rocchesso@unipa.it

**MARIA MANNONE**

University of Minnesota, alumna

manno012@umn.edu

*Embryo of a Quantum Vocal Theory of Sound*

Concepts and formalism from acoustics are often used to exemplify quantum mechanics. Conversely, quantum mechanics could be used to achieve a new perspective on acoustics, as shown by Gabor studies. Here, we focus in particular on the study of human voice, considered as a probe to investigate the world of sounds. We present a theoretical framework that is based on observables of vocal production, and on some *measurement apparati* that can be used both for analysis and synthesis. In analogy to the description of spin states of a particle, the quantum-mechanical formalism is used to describe the relations between the fundamental states associated with phonetic labels such as phonation, turbulence, and slow myoelastic vibrations. The intermingling of these states, and their temporal evolution, can still be interpreted in the Fourier/Gabor plane, and effective extractors can be implemented. This would constitute the basis for a Quantum Vocal Theory of sound, with implications in sound analysis and design.

**ANTONIO MASTOGIACOMO**

Conservatorio "N. Sala" Benevento

ant.mastrogiacomo@gmail.com

*On Mechanical Music starting with John Philip Sousa*

If the possibility of recording, listening again and manipulating sound has represented a delicate moment in the history of perception in the measured times of daily life, then the uncontrolled diffusion of the apparatuses - by which we mean the technical media of modernity, underlining their prepared and therefore intentional, artificial, conventional character - has been rooted in the wide horizon of the medium, a term already mentioned in the essay on technical reproducibility to name the way in which human perception is organized. And if at the root of media theory there is an idea according to which sensitive experience cannot be considered in superhistoric terms, then Sousa's words historicize this delicate moment in the measurement of time and of production and listening technologies. Starting from the essay *The Menace of Mechanical Music (1906)* by John Philip Sousa, the proposed paper discusses the function and the fruition of mechanical and mechanized music; referring to the experience of sound within the limits of mass society, five other different authors (Adorno, Benjamin, Edison, McLuhan and Moholy-Nagy) are placed in dialogue with the essay by the American composer and band conductor.

**FEDERICO AVANZINI, ADRIANO BARATÈ, GOFFREDO HAUS, LUCA ANDREA LUDOVICO, STAVROS NTALAMPIRAS, GIORGIO PRESTI**

LIM - Laboratorio di Informatica Musicale, Dipartimento di Informatica, Università degli Studi di Milano, Via Celoria 18 - 20133 Milano, Italia

lim@di.unimi.it

**DAVIDE ANDREA MAURO**

Department of Computer and Information Technology, Marshall University, 1548 6th Ave. Apt. 19 - Huntington, WV, 25701 (USA)

maurod@marshall.edu

*The Future of the IEEE 1599 Format*

The IEEE 1599 format is an international standard for multi-layer representation of music information. The Music Informatics Laboratory (Laboratorio di Informatica

Musicale - LIM) of the University of Milan is the working group that mostly contributed to the promulgation of the format and to the development of an ecosystem of applications for the generation, manipulation and fruition of information encoded according to this standard. Ten years after the official release of the IEEE 1599 specifications, in this work we investigate the critical points that have limited its diffusion so far, and we identify the new directions to be taken to make the format evolve and to contribute to its future success.

### **FABIO DE SANCTIS DE BENEDICTIS**

ISSM "P. Mascagni" - Livorno  
fabio.desanctis@consli.it

#### *FDSDB XXth CT: an Algorithmic Composition Library for Open Music and PWGL, Devoted to XXth Century Composition Techniques*

This paper describes FDSDB\_XXth\_CT, an algorithmic composition library, for PWGL and Open Music, devoted to some of XXth Century composition techniques. It can be freely downloaded at the following web page: <https://sites.google.com/site/fdsdbmascagni/code>. After a general review, necessarily not exhaustive, about the type of libraries at disposal for PWGL and Open Music, we move on to describe the process of realization that has led to the creation of the library, consequential to composition and analysis practice, and resulting by a work extended over the years, and still in progress. Afterwards the library structure is described in summary, and contemplated composition techniques are listed, afferent to Dodecaphony, Serial and Post-serial period, Pitch-Class Set Theory and other processes for developing pitches and durations. Finally some tutorial patches are showed, followed by a concise conclusion about aims and use of the library.

### **ANDREA VALLE**

StudiUm/CIRMA - Università di Torino  
andrea.valle@unito.it

#### *SampComp: Sample-based Techniques for Algorithmic Composition*

The paper introduces a series of techniques for algorithmic composition in which compositional data are extracted from audio at the sample level. In this way, typical Digital Signal Processing operations can act as tools to create/retrieve music materials. First, in order to describe audio-driven processes in composition, the notion of audioparity is introduced. Secondly, various approaches are discussed that map sample values to composition events. The implementation of these processes is then presented, in particular in relation to music notation generation. Finally, the paper discusses some applications of the same techniques to physical computing.

### **GIORGIO KLAUER**

Conservatorio "B. Marcello", Venezia  
giorgio@klauer.it

#### *The "Scuba" Prototype: an Augmented Garment to Design the Sound Interaction with Clothes*

A prototype of a garment designed as a tool for designing sound and sound interaction with clothes is presented. The salient features of the device (made of neoprene, sensors and speakers applied, built-in microcomputer, remote management system) give complex qualities in terms of *interaction design* and *embodiment*. Some experiences of use are described and, given the context in which the project takes place, some considerations are made about the positioning of this type of research.

### **MICHELE GERONAZZO, ROLF NORDAHL, STEFANIA SERAFIN**

Dept. of Architecture, Design & Media Tech., Aalborg University, Copenhagen  
fimg, rn, stsf@create.aau.dk

### **EDOARDO DEGLI INNOCENTI, DIEGO VESCOVI**

Dept. of Information Eng., University of Padova  
edosigma@gmail.com  
ves.diego@gmail.com

### **FEDERICO AVANZINI**

Dept. of Computer Science, University of Milano  
avanzini@di.unimi.it

#### *Technologies for Music Teaching: an Experience with Virtual Reality*

Mobile virtual reality is becoming popular and accessible to anyone with a recent smartphone. In this context digital didactics, starting from primary education, can exploit the natural interaction and immersion available in virtual environments. This article presents a technical and methodological framework to support the use of mobile virtual reality in primary school, aimed at improving music learning in terms of active listening, attention, and time. We present a multiplatform application named VR4EDU, and discuss the related design choices as well as the virtual reality scenarios developed, with specific reference to an innovative "virtual locomotion" solution that is found to be intuitive and usable. VR4EDU applications can run on low-cost devices, thus being an attractive option for didactic uses. The proposed approach has been tested on a study with two classes in the last year of a primary school, focused on music genre identification. A preliminary evaluation of the results shows an improvement in learning genre characterization compared to traditional lessons with printed materials and passive listening.

### **STEFANO DELLE MONACHE**

Dept. of Architecture and Arts, Iuav University of Venice  
sdellemonache@iuav.it

### **MICHELE BUCCOLI, LUCA COMANDUCCI, AUGUSTO SARTI**

DEIB, Politecnico of Milano University

name.surname@polimi.it

**GIOVANNI COSPITO, ENRICO PIETROCOLA, FILIPPO BERBENNI**

Conservatory of Music "G. Verdi" of Milano

name.surname@consmilano.it

*Time is Not On My Side: Network Latency, Presence and Performance in Remote Music Interaction*

A pilot test on the sense of presence and the quality of the interaction in Networked Music Performance is presented. Subjective measures, based on a presence questionnaire, are combined with objective quality metrics, in order to stress the contribution of temporal factors (i.e., network latency) on the musical experience in the mediated environment. Preliminary results in the scope of chamber music practice are presented.

**MARTINO SAROLLI**

Conservatorio "G. Verdi", Torino

**ANDREA AGOSTINI**

Conservatorio "G. Verdi", Torino

andrea.agostini@conservatoriotorino.eu

*Deep learning and digital sound processing: integration possibilities in electroacoustic compositional praxis*

The term *deep learning* generally refers to a repertoire of *machine learning* techniques based on multi-layer artificial neural networks. Key aspect and common ground is the term *deep*, which indicates the presence within the neural networks of multiple layers that process different levels of abstraction, making them capable of expressing complex representations in terms of simpler ones. Their ability to control high levels of complexity make *deep learning* systems able to represent and handle high-level abstraction data like the tonal-spectral and dynamic features of a sound. The possibility of accessing and operating on the internal representations of the *deep neural networks* allows the development of new processes of analysis, processing and sound synthesis. Among the first systems made available to experiment with these new possibilities, the Nsynth autoencoder will be presented and an attempt to integrate the techniques that it makes available inside the compositional praxis of an electroacoustic piece will be proposed.

**LAURA ZATTRA**

Conservatorio di Musica di Rovigo

STMS - IRCAM, CNRS

laura\_zattra@yahoo.it

**NICOLAS MISDARIIS**

STMS Ircam-CNRS-SU, Sorbonne Université

misdariis@ircam.fr

**FRANK PECQUET**

Pantheon Sorbonne University - CNRS

fpecquet@univ-paris1.fr

**NICOLAS DONIN**

STMS Ircam-CNRS-SU, Sorbonne Université

donin@ircam.fr

**DAVID FIERRO** CICM EA 1572, MUSIDANSE EDESTA - Paris8

davidfierro@gmail.com

*Analysis of Sound Design Practices. Research Methodology*

This contribution describes an on-going project started in March 2018 entitled "Analysis of Sound Design Practices" [ASDP]. The team comprises members from the LABEX CAP (Création, Arts et Patrimoines); the ACTE Institute (Arts Creation Theory Aesthetics); IRCAM's APM (Analysis of Musical Practices) and PDS (Sound Perception and Design) teams. By means of a large-scale study - based on web documentation, an online questionnaire, interviews and analytical results - the team is working towards understanding the creative process as well as the identity of Sound Design and Sound Designers in particular, in Europe. The main target concerns the development of a consistent sociological, geographical and historical knowledge of the discipline, and, consequently, to better identify the relevant issues in both artistic, technical or scientific commitment. The purposes of this article is to describe (a) the general framework of the project; (b) the methodology of the ongoing study (collaborative tools, writing and sending a questionnaire); (c) initial developments and results - terminology and definition, translation issues, conception of a database, listing of professionals and different institutions in Europe, first feedback analysis.

**SANDRA PAULETTO**

Department of Theatre, Film and TV, The University of York

sandra.pauletto@york.ac.uk

**BART WALUS**

Department of Theatre, Film and TV, The University of York

bartekwalus@hotmail.com

*Reflecting on the Role of Sound in an Immersive Multi-media Intervention for Health Communication*

We reflect on the role of sound in an immersive multimedia health communication intervention: *Jane's Story*. Health communication aims to inform, educate, and

ultimately produce behavioral change (sometimes urgently, sometimes longterm). *Jane's Story* is an immersive eight minutes animation aimed at young people. It is produced as a tool to be used by educators to stimulate discussions around the topic of chronic diseases and disorders (diabetes, obesity, cerebral palsy, etc.) in adolescents. The intervention design merges visual methods typically used in health communication (infographics, data visualisation), with storytelling techniques typically found in filmmaking (first-person voiceover, use of music and sound design). We surveyed 37 subjects on the effectiveness of the intervention and the perceived role of its soundtrack. The analysis of the results reveals that the intervention was perceived as being generally effective, and that voice, music and sound design strongly support the communication of the health message and the emotional engagement of the audience with the topic. We then consider these results through the lens of Construal Level Theory, which describes how psychologically distant and urgent an issue is perceived to be depending on how it is communicated. We then suggest that sound might be particularly effective in communicating concrete and urgent health messages that require a quick change, while symbolic visuals might be more effective portraying more general health messages, requiring longterm change. Additionally, this intervention shows that these two approaches can be combined effectively. Finally we report limitations and lessons learnt from this particular intervention that should be taken into account when designing multimedia health communication interventions.

### **GIOVANNI ALBINI**

Conservatorio "Jacopo Tomadini", Udine, Italy  
Eesti Muusika- ja Teatriakadeemia, Tallinn, Estonia  
mail@gioviannibini.it

#### *bach: A Procedural Video Game Featuring Music Visualization for an Interdisciplinary Computer Based Approach to the Mathemusal Education*

*bach*, a side-scroller video game developed by the author aiming to strengthen the interdisciplinary teaching of musical and mathematical knowledge, is introduced. Its gaming experience and the implications of its procedural level generation in the context of music visualization are then discussed in terms of their potential educational effectiveness with reference to experimentation results of a case study.

### **ALBERTO NOVELLO**

Conservatorio di Padova  
albynovello@gmail.com

### **ALESSANDRO FIORELMONDO**

Conservatorio di Padova  
dr.clipper@yahoo.it

### **ARCANGELO DI DONATO**

Conservatorio di Padova  
arcadido@gmail.com

#### *PhOS: a Photo-Sensitive Gestural Surface*

We present PhOS: a system that uses incident light intensity to control music parameters and sound synthesis. It is a transparent surface with 16 sensors each measuring ambient light, red, green and blue intensity values. The sensors are spaced on a 4x4 grid, making gestural control quite intuitive for the user, through the creation of light trajectories on the surface. Given its simplicity, with the direct translation of light into sound, and the ubiquitous distribution of light-emitting devices such as inexpensive bike lights and mobile phone flash lights, it can be used in multiple occasions from live performances to sound installations.

### **ARCANGELO DI DONATO**

Conservatorio "C. Pollini", Padova  
arcadido@gmail.com

### **GIORGIO KLAUER**

Conservatorio "B. Marcello", Venezia  
giorgio@klauer.it

#### *The Present Sound of Past Places*

The study presents an example of an interactive device that allows a differential experience of a current soundscape and an ancient map. Conceived as a book augmented through sonic and video interaction, the device allows to slide the finger over sensitive areas of a cartography of Renaissance Padua, revealing the current environmental sound along with images. The difference between the current soundscape and the representation of the urban landscape in the sixteenth century causes a discrepancy whose cultural value is placed in relation to acoustic ecology and the qualitative aspects of the experience to the field of urban planning.

### **ARCANGELO DI DONATO**

Conservatorio di Padova  
arcadido@gmail.com

### **ALBERTO NOVELLO**

Conservatorio di Padova  
albynovello@gmail.com

### **ALESSANDRO FIORELMONDO**

Conservatorio di Padova  
dr.clipper@yahoo.it

*The Pictorial Controller: a Gestural Instrument for Sound Synthesis*

The “pictorial controller” allows to control the sound synthesis by painting. The instrument consists of a table whose worktop is equipped with sensors that communicate with the music software the chromatic choice and the gestures of the artist through the electrical conductivity of water and graphite. Colors and signs are translated musically in real time.

**ALESSANDRO BARILE**

Conservatorio “G. Tartini” Trieste  
barilealessandro3@gmail.com

**SAMUEL MONTALTO**

Conservatorio “G. Tartini” Trieste  
montalto.samuel@gmail.com

**LUCA CIGAINA**

Conservatorio “G. Tartini” Trieste  
luccacigaina@gmail.com

*Safe and Sound*

With “Safe and Sound”, the English version of our way of saying “safe and sound”, we are referring to a system that aims to make new generation appliances safer by providing sound monitoring of the operational status of the appliance and its interaction with the user. The technology has focused on the production of increasingly silent instruments in their operating mechanisms and because of this process have been eliminated information related to the acoustic, natural, mechanical sound of the instrument itself. The technology can however be used to implement an audio monitoring system of the behavior of the instrument capable of restoring usage information. In this work we have applied the concept of “Safe and Sound” to an induction cooktop and an iron.

**LORENA CANTARUT**

Conservatorio “G. Tartini” Trieste  
lorena.cantarut@gmx.com

*LIS-10 and Heed: two Applications of Shape Memory Alloys in Sound Art*

This article presents the application of a Shape Memory Alloy (SMA) as part of Sound Art. The main property of an SMA is its ability to recall a specific shape previously imprinted. In the present work, an SMA consisting of nickel and titanium, called Nitinol, was used. The crystalline structure of Nitinol can assume two different phases that depend on the heating of the alloy: at room temperature it is malleable, and therefore deformable, while, if heated above a certain temperature, it is able to reorganize its crystalline structure so as to assume a predetermined shape. This phase shift is what makes SMAs ideal for human-machine interaction. In this case the temperature is controlled through a circuit and digital algorithms. In particular, I have chosen to use sound analysis as an algorithm. In other words, the sound sculpture is changed shape by means of sound descriptors, extracted from the acoustic input produced by the users of the installation.

**MICHELA COPPOLA**

Conservatorio di Musica “Giuseppe Verdi” Milano  
micopp77@yahoo.it

*MYTRENDS: Playing Words*

MyTrends is an interactive ‘musical machine’ as it’s capable of generating proper musical compositions. The aim of the study is to give everyone the ability to create a piece of music, whether or not the subject involved has the required skills for composing music. This is achieved with an intuitive and simple tool. The verb ‘to play’ in English language, as for the French ‘jouer’, is well referred to the word ‘game’ as to ‘music’: indeed at all effects, this project is inspired to musical customs derived from ancient pastimes mainly dating back to XVIII and XIX centuries in Europe, one example of which could be some musical dice games. To convey a more intuitive and fast approach for the user, it has been chosen the widespread tool that is the internet today. MyTrends asks the user to carry out some day-to-day operations, as sitting in front a pc keyboard and a monitor - on which the well familiar Google banner appears - and to make a research for a single word. MyTrends interviews Google Trends report system and interprets the output data, while running the sonification algorithm linked to the word that has been written down by the user. The system allows the user to take part to the interpretation steps and to include some variations of the sonic material that has been generated by the machine, via a mouse interaction.

**KRISTJAN STOPAR**

Università degli Studi di Udine  
kristjan.stopar@uniud.it

**ROBERTO CALABRETTO**

Università degli Studi di Udine  
roberto.calabretto@uniud.it

**LUCA COSSETTINI**

Università degli Studi di Udine  
luca.cossettini@uniud.it

*New Music Writing Processes for Cinema. Recording Studios, Digital Media and Compositional Practices*

The new music writing processes for cinema are increasingly characterized by software, which allow composing, mixing, and dubbing music. This fact has made the traditional composition practices obsolete and relegated them to the fringes of film music production. As Hans Florian Zimmer pointed at, the computer has nowadays

become an essential musical instrument, “and then more virtuosistic you get on that, the better you can express your ideas”. However, despite the fact that a considerable number of composers use the computer as a real compositional instrument, a scientific in-depth study of basic techniques of the writing method and technological procedures has not been conducted yet. Most musicians, supported by this new computer environment, have moved the focus of the creative act from traditional compositional practice to the universe of sound design, as confirmed by countless interviews with Hollywood composers, including Hans Zimmer and Tom Holkenborg. Therefore, it is increasingly necessary to develop a scientific study, able to combine music composition with the use of software and apply it to moving images. Thereby it will be possible to create the conditions for a scientific debate and the following production of film music for new media.

### **LUCA RICHELLI**

Sound and Music Processing Lab, Conservatorio “C. Pollini” - Padova  
lucarichelli@gmail.com

#### *Motion Capture and Augmented Instruments: Gestural Control and Compositional Gesture*

This paper presents the research and production carried out at the Sound and Music Processing Lab (SaMPL) of the Padua Conservatory in collaboration with the Centro di Sonologia Computazionale (CSC) of the Department of Information Engineering (DEI) of the University of Padua, with the PhaseSpace motion capture system. It will be shown the technical characteristics of the system, the musical productions, the technical and musical problems inherent to the work done and the future prospects. The main area of the system has been the integration of the motion capture parameters of the instrumental performers with the music itself, or seen from a diametrically opposite point of view, the creation of hyper-instruments (or augmented instruments) that allow the performer to control, besides the acoustic emission of the traditional instrument, new possibilities for electroacoustic expressions. The executive gesture integrates the acoustic and electronic dimension of the musical work becoming an integral part of the score, requiring the composer to think of this new dimension too.

### **YURI DE PRA**

University of Udine, Department of Mathematics, Computer Science and Physics, Via delle Scienze 206, Udine 33100, Italy  
depra.yuri@spes.uniud.it

### **FEDERICO FONTANA**

University of Udine, Department of Mathematics, Computer Science and Physics, Via delle Scienze 206, Udine 33100, Italy  
federico.fontana@uniud.it

### **MICHELE SIMONATO**

Electrolux Professional S.p.A., Viale Treviso, 40, Pordenone 33170, Italy  
michele.simonato@electrolux.it

#### *Development of Real-Time Audio Applications Using Python*

The development of real-time audio applications for general purpose OS like Linux or MS Windows is a well-known and not trivial problem. Several real-time audio software with libraries for multi-purposes programming languages like C, C++ or Java have been developed. Although these solutions provide portability through multiple OS, they often introduce problems related to dependencies and versions alignment over the time. In the present article, the process to build a framework for real-time audio relying on the programming language Python is described. We selected Python because it offers a fast learning rate coupled with several modules that can be used inside real-time audio projects. The advantages of this programming language but also the limitations in terms of execution time imposed by the interpreted code are both discussed. In particular, a procedure of code refactoring to obtain the real-time execution is presented. This procedure involves the inclusion of compiled code generated by Cython and C language, only where strictly needed. In the last part, some case studies related to the implementation of Virtual Analog models are presented introducing different amounts of compiled code and measuring the execution time.

### **ARCANGELO DI DONATO**

Conservatorio “C. Pollini”, Padova  
arcadido@gmail.com

### **ALBERTO NOVELLO**

Conservatorio “C. Pollini”, Padova  
albynovello@gmail.com

### **GIORGIO KLAUER**

Conservatorio “B. Marcello”, Venezia  
giorgio@klauer.it

#### *The Electroacoustic Jewel*

The Electroacoustic Jewel is a small wearable electronic circuit that emits or is operated by sound. In this article, we introduce a field of research and development for such jewels: beyond the pure display of aesthetics, these objects aim at adding interactivity to extend their fruition to the sense of hearing.

### **STEFANO CATENA**

Conservatorio “Giuseppe Verdi”, Milano  
stefano.catena23@gmail.com

### **ANDREA BOLZONI**

Conservatorio “Giuseppe Verdi”, Milano  
andreabolzoni@music@gmail.com

### *Distorsione di massa*

“Distorsione di Massa” is an interactive installation that, in the form of a game, will allow the user to think about the fake news topic. It will appear as a listening position, made of a chair and a table on which a tablet is found, positioned in the center of a quadrasonic audio diffusion space. The game is organized in levels of increasing difficulty, where the objective is to recognize a musical phrase representing authentic news. At the beginning of each level the “authentic” news will be reproduced, later multiplied and then passed through various distortion processes. After a certain amount of time proportional to the level’s difficulty, the possibility of choosing on the tablet which of the sources has remained unaltered, will be unlocked.

### **GIOVANNI SPARANO**

Ricercatore indipendente  
giovannsparano@gmail.com

### *GrainLab - Open Source Software for Quasi-Synchronous Granular Synthesis*

This paper presents GrainLab, an open source software for granular synthesis in real time. In its implementation, specific attention was paid to scalability, usability and efficiency, providing access to the most parameters, a simple and accessible graphic interface and optimizing the grain generation algorithm. The software is free and open source, therefore it can be easily edited and adapted to specific needs.

### **PAOLO ZAVAGNA**

Conservatorio “B. Marcello”, Venezia  
paolo@zavagna.it

### **FRANCESCO PECCOLO**

Conservatorio “B. Marcello”, Venezia  
francescopeccolo96@gmail.com

### **RICCARDO SELLAN**

Conservatorio “B. Marcello”, Venezia  
riccardo.sellan@hotmail.it

### *Electroacoustic and Computer Music Timelines*

Electroacoustic and Computer Music timelines are projects created by Paolo Zavagna in collaboration with students from Conservatories of Venezia and Padova (Italy). Timelines are organized by events, biographies, compositions, technologies and other relevant topics about electroacoustic and computer music, each one viewable with additional information. The project is totally open-source and uses the SIMILE widget (MIT) that adapts to the most commonly-used browsers.


### **GIACOMO LEPRI**

Centre for digital music - Queen Mary University of London  
g.lepri@qmul.ac.uk

### *InMuSIC: an Interactive Multimodal System for Electroacoustic Improvisation*

InMuSIC is an Interactive Musical System (IMS) designed for electroacoustic improvisation (clarinet and live electronics). The system relies on a set of musical interactions based on the multimodal analysis of the instrumentalist’s behavior (embodied motion qualities and sonic production). The procedures outlined for the real-time organisation of the electroacoustic materials intend to facilitate the shared development of musical structures. Aim of the project is to investigate compositional and performative strategies for the establishment of a musical collaboration between the improviser and the system.





Contributi  
*Papers*



# Aptica e interazione musicale

Stefano Papetti

Institute for Computer Music and Sound Technology  
 Zürcher Hochschule der Künste  
 stefano.papetti@zhdk.ch

## ABSTRACT

Durante la pratica musicale, tra musicisti e strumenti tradizionali si instaura tipicamente una complessa interazione aptica (tattile e propriocettiva) che è di sostegno all'esecuzione sia sotto il profilo manuale (per esempio facilitando la precisione temporale, di intonazione e dinamica) che estetico/espressivo. Al contrario, l'esperienza fisica attualmente offerta dall'elettronica di consumo è estremamente elementare se confrontata con la ricchezza della nostra interazione quotidiana con l'ambiente circostante. In particolare, il miglioramento dell'interazione aptica con dispositivi elettronici per l'interazione musicale, quali controller e strumenti musicali digitali, richiede ancora un sostanziale contributo da parte della ricerca in ambito tecnologico, percettivo e di interazione uomo-macchina. Questo lavoro discute la recente attività dell'autore in tale scenario di ricerca.

## 1. INTRODUZIONE

Il termine *haptics* ("aptica" in italiano) definisce un settore tecnico-scientifico interdisciplinare che verte sul sistema somatosensoriale – in particolare senso del tatto e proprioccezione – studiandone le caratteristiche percettive, fisiologiche e biomeccaniche, e sviluppando dispositivi che stimolino e si interfaccino con questi canali sensoriali.

Le interfacce aptiche vengono attualmente studiate e sviluppate sia a livello industriale *high-tech* che *consumer*, con applicazioni come teleoperazione, controlli per piloti, protesi e riabilitazione, realtà virtuale e intrattenimento. Tale interesse trova la sua motivazione nella convinzione che il feedback aptico possa essere utilizzato efficientemente come canale di comunicazione alternativo nell'ambito dell'interazione uomo-macchina permettendo di ridurre il carico sensoriale e cognitivo legato alla visione [1], al contempo risultando in maggiore naturalezza, coinvolgimento e realismo per l'utente.

Esistono diversi punti di contatto tra il campo dell'aptica e quello musicale. Per cominciare, il suono non è altro che la manifestazione acustica di un processo vibratorio. Quando partecipiamo ad un concerto veniamo investiti da un fronte di onde sonore più o meno intense che producono sia stimoli uditivi che tattili, e ulteriori stimoli tattili si

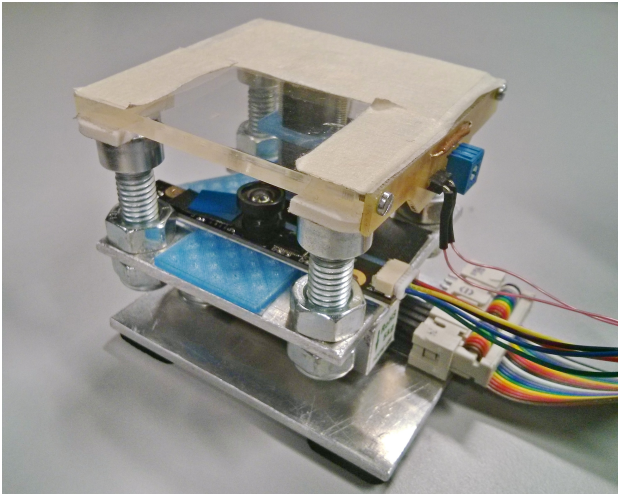
propagano attraverso oggetti solidi in contatto con il nostro corpo quali poltrone/seggiole e pavimentazione. Allo stesso tempo, i musicisti sul palco sono soggetti ad un complesso insieme di stimoli audio-aptici provenienti dai loro strumenti e da eventuali sistemi di amplificazione. Gli aspetti uditivi e acustici dello scenario musicale sono da tempo oggetto di ricerca interdisciplinare, mentre la prospettiva aptica è stata presa in considerazione relativamente di recente [2]. Gli obiettivi principali di una "aptica musicale" (*musical haptics*) sono da un lato quello di definire il ruolo dell'aptica nell'ascolto e nella pratica musicale, e dall'altro quello di sviluppare dispositivi musicali che offrano feedback aptico avanzato. La prima linea di ricerca richiede l'implementazione di vari esperimenti che spaziano dalla percezione audio-aptica fino a test prestazionali (dati oggettivi: per esempio, l'accuratezza nell'eseguire un dato task musicale) e valutazioni qualitative (dati soggettivi: per esempio, la qualità percepita di uno strumento). La seconda linea di ricerca è principalmente di carattere tecnologico e progettuale, richiedendo lo sviluppo e/o l'ottimizzazione di soluzioni tecnologiche per rendere stimoli aptici, sia a livello hardware (per esempio, attuatori vibrotattili) che software (per esempio, tecniche di sintesi di segnali aptici come forze e vibrazioni). Idealmente quest'ultima linea di ricerca (tecnica) si dovrebbe basare su risultati derivanti dalla prima (ricerca sperimentale). Al contempo l'implementazione di prototipi di dispositivi musicali aptici è necessaria per la realizzazione di esperimenti atti a testarne l'efficacia.

L'ipotesi principale dell'aptica musicale è che un feedback aptico avanzato generato da dispositivi elettronici musicali permetta di offrire una migliore e più completa esperienza per l'utente: per il musicista, migliorandone le prestazioni e l'espressività (ad esempio ricreando il naturale scambio fisico che si instaura con gli strumenti tradizionali), e per musicisti ed ascoltatori, migliorando la qualità dell'interazione e/o dell'ascolto.

Ulteriori potenziali applicazioni dei dispositivi musicali aptici si trovano nell'ambito dell'insegnamento, come guida durante l'esecuzione (specialmente nel caso di musica elettronica o performance remote) e in ambienti intrinsecamente rumorosi come grandi ensemble, e per facilitare l'accesso alla pratica musicale da parte di persone differentemente abili o affette da disordini somatosensoriali.

Nel seguito di questo articolo vengono descritti i principali risultati di ricerca sperimentale dell'autore e collaboratori nell'ambito dell'aptica musicale.

Copyright: ©2018 Stefano Papetti. This is an open-access article distributed under the terms of the [Creative Commons Attribution License 3.0 Unported](https://creativecommons.org/licenses/by/3.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.



**Figura 1.** L'interfaccia aptica sviluppata e utilizzata ai fini dell'esperimento [6].

## 2. ESPERIMENTI

In questa sezione sono riportati alcuni recenti esperimenti condotti dall'autore e collaboratori, atti a verificare come somatosensazione e udito possano influenzare vari aspetti dell'interazione gestuale e musicale.

### 2.1 Sensibilità alle vibrazioni in condizioni di tocco attivo

La sensazione tattile è stata oggetto di ampio studio nel secolo scorso [3, 4], individuando le cellule cutanee (meccanorecettori) incaricate di trasdurre segnali vibratorii, e trovando, tra le altre cose, che la sensibilità cutanea è massima alle mani, labbra e piedi, e dipende dal contenuto in frequenza degli stimoli vibratorii. In particolare per le dita della mano, risultano percepibili vibrazioni nell'intervallo di frequenze tra i 40 e 1000 Hz, con un picco di sensibilità tra i 200 e i 300 Hz. Tale intervallo, corrispondente approssimativamente all'intervallo tra le note E1 e B5, si sovrappone con l'intervallo di frequenze udibili, e ancora di più con frequenze di interesse musicale.

Gli studi sopra citati, classicamente consideravano il tatto in condizioni passive, ossia richiedendo ai soggetti di rimanere immobili (ad esempio assicurando il dito oggetto di studio ad un supporto), e stimolando la cute tramite dei dispositivi meccanici appositamente realizzati (ad esempio un puntale con movimento unidirezionale e area di contatto predefinita) [5].

Lo studio che viene qui riportato esamina invece la sensibilità alla vibrazione in un contesto più affine all'esperienza quotidiana, per cui è tipicamente attraverso l'azione di toccare qualche cosa che entriamo in contatto con stimoli vibratorii. Questo è appunto quanto accade tipicamente quando si suona uno strumento musicale tradizionale come archi, pianoforte, chitarra, ecc.

#### 2.1.1 Descrizione dell'esperimento

L'esperimento richiedeva ai soggetti di premere il dito indice della mano dominante su di una superficie piatta (pannello superiore dell'interfaccia mostrata in Figura 1) fino a

|       | Onda sinusoidale (N=27) |                      | Rumore filtrato (N=17) |
|-------|-------------------------|----------------------|------------------------|
| Forza | Accelerazione (rms)     | Dislocamento (picco) | Accelerazione (rms)    |
| 1,9 N | 76.1 dB                 | -48.7 dB             | 83.1 dB                |
| 8 N   | 69.7 dB                 | -58.2 dB             | 72.6 dB                |
| 15 N  | 68.5 dB                 | -59.3 dB             | 75.4 dB                |

**Tabella 1.** Percezione di vibrazioni in condizioni di tocco attivo. Soglie medie espresse come accelerazione RMS (riferimento dB:  $10^{-6} \text{ m/s}^2$ ) al variare della forza applicata. Il dislocamento massimo corrispondente viene riportato per lo stimolo sinusoidale (riferimento dB:  $10^{-6} \text{ m}$ ).

raggiungere un livello prestabilito di forza (tre target: 1,9 / 8 / 15 N). A quel punto, seguendo il protocollo sperimentale *two-alternative forced choice* (2AFC) venivano considerati due intervalli temporali di circa 1,5 secondi ciascuno, rispettivamente nominati A e B, e in uno solo di essi veniva generato uno stimolo vibratorio attraverso la superficie in contatto con il dito. I soggetti dovevano quindi riportare allo sperimentatore in quale dei due intervalli avesse percepito la vibrazione. L'intensità della vibrazione era inizialmente sufficientemente elevata da essere chiaramente percepita, e con il proseguire dell'esperimento poteva essere diminuita o incrementata di 2 dB secondo il protocollo sperimentale *one-up two-down* [7]: un incremento per singola risposta negativa (vibrazione non percepita) e decremento per due risposte positive (vibrazione percepita) consecutive.

Come stimoli vibrotattili si sono utilizzati un'onda sinusoidale a 250 Hz e rumore bianco filtrato nell'intervallo 40-500 Hz (filtro passabanda). Questa scelta è stata compiuta per massimizzare la percezione delle vibrazioni, sulla base dei risultati della precedente letteratura menzionati all'inizio della sezione. Per evitare che i soggetti potessero sentire il suono emesso dall'attuatore di vibrazione al suo funzionamento – così facilitando la scoperta del caso in cui la vibrazione era presente – è stato chiesto loro di indossare delle cuffie con del rumore di mascheramento durante lo svolgimento dell'esperimento.

Ventisette soggetti hanno partecipato all'esperimento con lo stimolo sinusoidale e diciassette con lo stimolo rumoroso. L'età variava entro 19-39 anni (media=26, sd=4.5) e metà dei soggetti erano studenti di uno strumento musicale a livello universitario.

#### 2.1.2 Risultati

I risultati dello studio, riportati in Tabella 1, hanno dimostrato che le soglie percettive vibrotattili nelle condizioni dell'esperimento sono significativamente inferiori a quelle riportate nella letteratura precedente. Inoltre, le soglie vanno generalmente abbassandosi all'aumentare della forza applicata sulla superficie vibrante. Più precisamente, le soglie misurate nell'esperimento come accelerazione RMS variano tra 68.5 e 83.1 dB (riferimento dB:  $10^{-6} \text{ m/s}^2$ ) in relazione alla forza applicata e alla natura dello stimolo, mentre le soglie riportate in letteratura per stimoli sinusoi-

dali tra 100 e 250 Hz variano tipicamente nell'intervallo 105-115 dB.

La motivazione principale alla base di questo fenomeno può essere attribuita alla presenza di "tocco attivo", per cui i soggetti premevano un dito sulla superficie dell'interfaccia sperimentale. Questo a sua volta implica i seguenti fattori che esercitano probabilmente effetti positivi sulle soglie: a) intenzione e azione; b) forza esercitata perpendicolarmente alla superficie di contatto relativamente elevata rispetto alla letteratura precedente. c) area di contatto relativamente elevata rispetto alla letteratura precedente. In particolare, la combinazione dei fattori forza e area porta un accoppiamento meccanico ottimale tra il polpastrello e la superficie vibrante e al contempo un maggiore numero di meccanorecettori interessati, con conseguente superiore trasduzione di stimoli vibratorii.

## 2.2 Influenza del feedback vibrotattile sulla valutazione del pianoforte acustico

La pratica musicale con strumenti tradizionali è un'esperienza largamente multisensoriale (uditiva, aptica e in parte anche visiva) in cui strumenti e musicisti sono solidamente accoppiati in un complesso sistema di interazioni [8]: lo strumento risponde ai gesti del musicista, e questo a sua volta calibra i propri gesti in base alla risposta sonora e aptica (vibrotattile e propriocettiva) dello strumento. Come confermato da alcuni recenti studi, principalmente concentrati su violino e pianoforte [9], risulta chiaro che la percezione e la valutazione della qualità di uno strumento non sono legate semplicemente alle sue caratteristiche sonore ma anche a quelle aptiche.

Gli esperimenti descritti in questa sezione hanno esaminato come la percezione di varie caratteristiche qualitative del pianoforte acustico possa essere influenzata dalla risposta vibrotattile della tastiera [10].

### 2.2.1 Descrizione dell'esperimento

Per l'esperimento si sono utilizzati due pianoforti Yamaha Disklavier, uno verticale ed uno a coda, i quali permettono la registrazione e riproduzione di partiture MIDI per mezzo di sensori e attuatori elettromeccanici incorporati. Tali pianoforti possono inoltre essere utilizzati in modalità silente, nella quale i martelletti non colpiscono le corde (non producendo quindi né vibrazione né suono) e trasformando di fatto lo strumento in una tastiera MIDI muta. Le modalità normale e silente sono state alternate durante lo svolgimento dell'esperimento, permettendo di esporre i soggetti pianisti rispettivamente alle vibrazioni naturali prodotte dallo strumento e ad una condizione senza vibrazioni.

Per il feedback sonoro si è utilizzato il sintetizzatore software Modartt Pianoteq, impostato per simulare un piano verticale o uno a coda a seconda del pianoforte effettivamente suonato dai soggetti, e pilotato dai dati MIDI (messaggi "note on", valori di velocity, e "note off") forniti dal Disklavier. L'audio è stato riprodotto attraverso cuffie isolanti o auricolari in-ear sopra cui erano indossate delle protezioni auricolari: l'obiettivo era quello mascherare quanto più possibile il suono acustico prodotto dal Disklavier nella modalità normale, e fare in modo che il feedback

sonoro in entrambe le modalità fosse sostanzialmente solo quello del sintetizzatore riprodotto via cuffia (cosa confermata in una fase pilota). Il volume e la risposta dinamica del sintetizzatore di pianoforte sono stati preliminarmente calibrati di modo da riprodurre l'intensità sonora originale del pianoforte acustico utilizzato.

L'esperimento richiedeva ai soggetti di utilizzare solamente le mani: i pedali sono stati resi inaccessibili, mentre la panca, i piedi dei pianisti e gli appoggi degli strumenti sono stati isolati dal pavimento tramite pannelli isolanti. Le misure di vibrazione effettuate hanno confermato che in questo modo l'intensità delle vibrazioni trasmesse dal pavimento non superava quella del rumore di fondo della stanza.

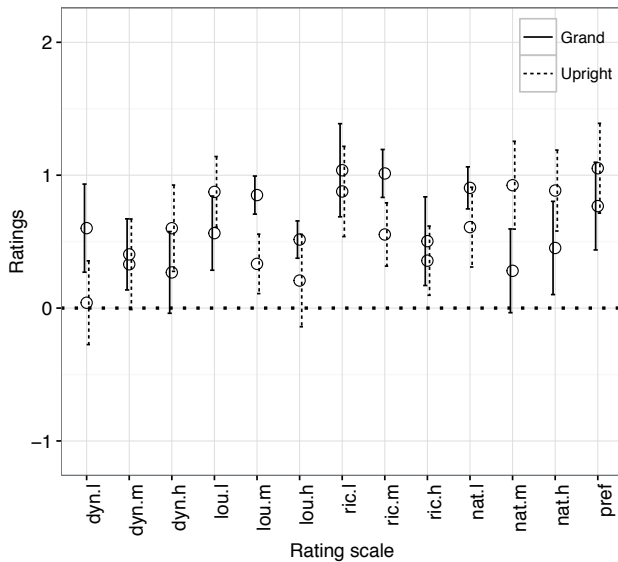
Un totale di 25 studenti di pianoforte a livello universitario (mediamente con 15 anni di esperienza) hanno partecipato all'esperimento, 15 sul piano verticale e 10 sul piano a coda. L'età media era 27 anni.

Il compito dei soggetti era quello di suonare liberamente (senza però utilizzare i pedali) e di confrontare due configurazioni A/B commutabili tramite un interruttore. I pianisti tuttavia non sapevano che cosa potesse cambiare tra le due configurazioni: una configurazione poneva il pianoforte Disklavier in modalità silente (assenza di vibrazioni), mentre l'altra corrispondeva alla modalità normale (vibrazioni naturali). L'assegnazione delle etichette A e B alle modalità con/senza vibrazioni è avvenuta in modo casuale. Il confronto consisteva nell'assegnare un punteggio a vari attributi qualitativi su una scala di valori continui tra -3 e +3, rispettivamente corrispondenti a "A è molto migliore di B" e "B è molto migliore di A". Le caratteristiche da valutare erano: dinamica, intensità sonora, ricchezza sonora, naturalezza e preferenza generale. Con l'eccezione della preferenza generale, tutti gli attributi sono stati valutati separatamente per tre settori distinti della tastiera, rispettivamente nominati "low" (tasti fino al RE della quarta ottava), "mid" (tasti tra il RE della quarta ottava e il LA della sesta), e "high" (tasti oltre il LA della sesta ottava).

### 2.2.2 Risultati

Come mostrato in Figura 2, la configurazione con vibrazioni naturali è stata generalmente preferita sia per il piano a coda che per quello verticale. Va tuttavia notato come i bassi valori medi di preferenza siano dovuti ad alcune marcate incoerenze tra le preferenze dei soggetti: 5 soggetti hanno generalmente preferito la configurazione senza vibrazioni.

Curiosamente, nella fase di rapporto post-esperimento solo un soggetto si è reso conto che l'unica differenza tra le due configurazioni stava nella presenza o meno di vibrazioni. Di interesse è anche il fatto che la grande maggioranza dei soggetti abbia chiaramente associato gli attributi di dinamica e intensità sonora – che non variano effettivamente, ma sono stati evidentemente influenzati dalla presenza di vibrazioni – con quelli di naturalezza, ricchezza e preferenza generale.



**Figura 2.** Valutazione della qualità del piano acustico in base alla presenza di vibrazioni. Valori positivi significano una preferenza per la configurazione con vibrazioni. Le etichette sull'asse x riportano delle abbreviazioni per le caratteristiche qualitative valutate (dyn = dinamica, lou = intensità sonora, rich = ricchezza sonora, nat = naturalezza, and pref = preferenza) e il settore di tastiera considerato (l = low, m = medium, h = high).

### 2.3 Percezione di vibrazione alla tastiera del pianoforte acustico

Il pianoforte presenta un'interfaccia ben definita e limitata con il musicista (mani sulla tastiera e piedi eventualmente sui pedali), il che rende relativamente agevole uno studio sistematico delle sue caratteristiche vibratorie durante la pratica strumentale. Vari studi hanno studiato le vibrazioni del pianoforte in un contesto di acustica musicale, ad esempio con misure alle corde e alla tavola armonica [11]. Solo di recente sono state invece studiate le vibrazioni così come trasmesse al pianista [12]. Un lavoro di riferimento in questo senso è quello di Askenfelt e Jansson [13] i quali hanno effettuato misure di vibrazione in varie posizioni sul corpo dei principali strumenti tradizionali a corde, tra cui il pianoforte. Di notevole interesse per la presente discussione è il fatto che, confrontando le soglie vibrotattili classiche in letteratura (tocco passivo) con l'ampiezza delle vibrazioni registrate alla tastiera del pianoforte, se ne trae che il pianista non dovrebbe essere in grado di percepire vibrazioni se non per multiple note suonate contemporaneamente con dinamiche di tipo fortissimo in registro grave. Lo studio di seguito descritto ha prodotto evidenze opposte a quanto descritto da Askenfelt e Jansson [10].

#### 2.3.1 Descrizione dell'esperimento

L'obiettivo dell'esperimento era di verificare la possibilità di percepire vibrazioni attraverso la tastiera del pianoforte durante la pratica strumentale. Si è utilizzata la stessa strumentazione descritta nella Sezione 2.2: questo ha permesso di controllare la presenza di feedback vibro-

tattile (modalità silente/normale), mantenendo il feedback sonoro inalterato.

Un totale di 28 studenti di pianoforte a livello universitario hanno partecipato all'esperimento, 14 sul piano verticale e altrettanti sul piano a coda. L'età media era 27 anni e mediamente avevano 15 anni di esperienza.

La percezione delle vibrazioni è stata misurata per singole e multiple note di varie altezze, durata e dinamica. Nello specifico, come note singole si sono considerati i tasti LA dalla prima alla sesta ottava (A0-A5 secondo lo standard MIDI) con durate "lunga" (8 battiti a 120 BPM) o "corta" (2 battiti), e dinamiche "forte" (da *mf* a *ff*, corrispondente a MIDI velocity da 72 a 108) o "piano" (da *p* a *mp*, corrispondente a MIDI velocity da 36 a 54). Inoltre si sono considerati i tasti RE della quinta e sesta ottava (D4 e D5) per durata lunga e dinamica elevata. In aggiunta alle note singole, si sono considerate delle triadi composte dalle note DO-RE-MI alla quinta e sesta ottava (CDE4 e CDE5 secondo lo standard MIDI). Lo scopo era quello di rinforzare il feedback vibratorio nelle zone della tastiera in cui si registra un evidente decremento della percezione, come provato in un precedente studio preliminare [14].

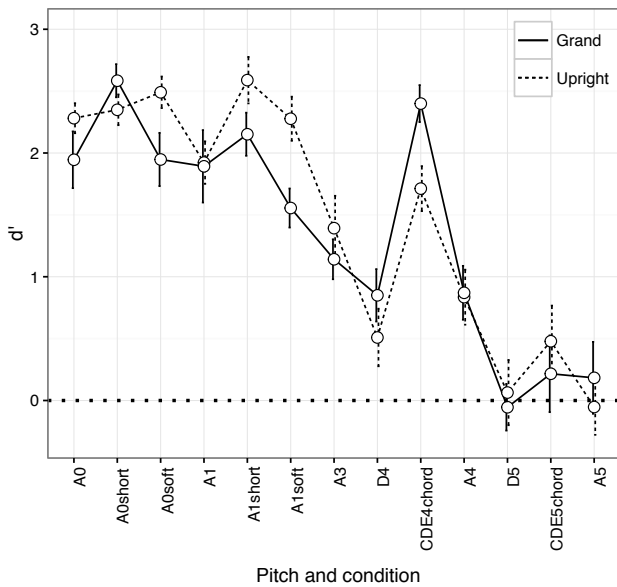
Ogni sessione sperimentale si svolgeva in due parti: nella parte A, onde massimizzare la percezione delle vibrazioni, i soggetti suonavano solo singole note lunghe e con dinamica forte; la parte B invece rendeva il compito più difficile nelle ottave basse, considerando singole note brevi e con dinamiche poco intense, mentre aveva l'intento di facilitare la percezione nelle ottave alte, considerando anche triadi di lunga durata e dinamica elevata.

Come nell'esperimento descritto nella Sezione 2.1, anche in questo si è adottata una procedura 2AFC che richiedeva ai soggetti di riportare se avessero percepito le vibrazioni mentre eseguivano i compiti descritti sopra. Ogni compito è stato ripetuto in ordine casuale ponendo il Disklavier 8 volte in ciascuna delle due modalità normale e silente, rispettivamente corrispondenti a vibrazione presente e assente.

L'esperimento è stato condotto sotto controllo di uno sperimentatore con l'ausilio di software sviluppato in Pure Data il quale si occupava di: a) leggere una scaletta descrittiva i compiti da eseguire e la condizione di feedback vibratorio; b) impostare i Disklavier nella modalità normale o silente; c) monitorare la corretta esecuzione del compito richiesto (note, durata e dinamica); d) registrare le risposte dei soggetti (si/no).

#### 2.3.2 Risultati

Per ogni soggetto e condizione si è calcolato un *indice di sensibilità*  $d'$  [15] come funzione della proporzione di risposte "si" corrette (ossia con vibrazione presente) e falsi allarmi (con vibrazione assente). I risultati dell'esperimento sono visualizzati in Figura 3. La pura casualità (probabilità del 50%) è rappresentata da  $d' = 0$ . Si può notare un chiaro andamento per cui la percezione delle vibrazioni diminuisce al crescere dell'altezza delle note. Al contrario, l'effetto delle condizioni di durata e dinamica non è particolarmente rilevante. Si riscontrano invece differenze evidenti tra la triade DO-RE-MI alla quinta ottava (CDE4)



**Figura 3.** Percezione di vibrazioni alla tastiera del piano acustico. Indice di sensibilità  $d'$  al variare delle condizioni sperimentali. Ove non specificato altrimenti, i tasti sono stati suonati con durata lunga e dinamica elevata.

e la relativa nota singola RE (D4), mentre la triade all’ottava superiore non presenta effetti significativi, suggerendo che alla quinta ottava lo stimolo su multiple dita abbia fatto recuperare sensibilità, mentre alla sesta ottava l’energia delle vibrazioni sia concentrata in una regione spettrale troppo elevata per essere percepibile [5].

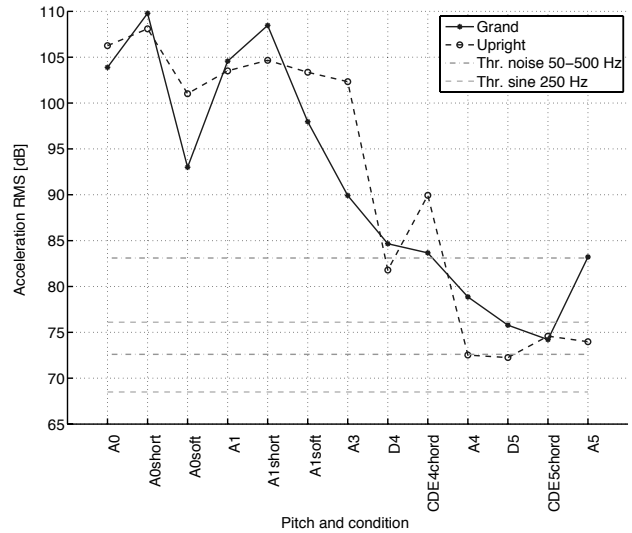
Come anticipato all’inizio di questa sezione, l’elevata percepibilità delle vibrazioni alla tastiera del pianoforte (soprattutto nelle prime quattro ottave) misurata dall’esperimento non può essere spiegata a livello psicofisico confrontando l’intensità delle relative vibrazioni alle soglie vibrotattili riportate nella letteratura classica [13, 5]. Al contrario, come mostrato in Figura 4, utilizzando le soglie per tocco attivo la cui misura è descritta nella Sezione 2.1 si ottiene una descrizione generalmente accurata del fenomeno percettivo registrato.

**2.4 Influenza del feedback vibrotattile sulla valutazione di pianoforte digitale**

Similmente a quanto riportato nella Sezione 2.2, l’esperimento descritto qui di seguito ha studiato come la valutazione qualitativa di un piano digitale possa essere influenzata da vari tipi di feedback vibratorio generati alla sua tastiera [16].

**2.4.1 Descrizione dell’esperimento**

Per l’esperimento si è modificato un piano digitale fissando un paio di potenti trasduttori vibrotattili al pannello inferiore dello strumento, così da poter generare vibrazioni arbitrarie alla tastiera (Figura 5). I trasduttori erano pilotati da un amplificatore collegato ad una interfaccia audio, e i segnali di vibrazione sono stati riprodotti mediante un campionario. Le condizioni di feedback vibrotattile considerate erano le seguenti: A) vibrazione assente; B) vi-

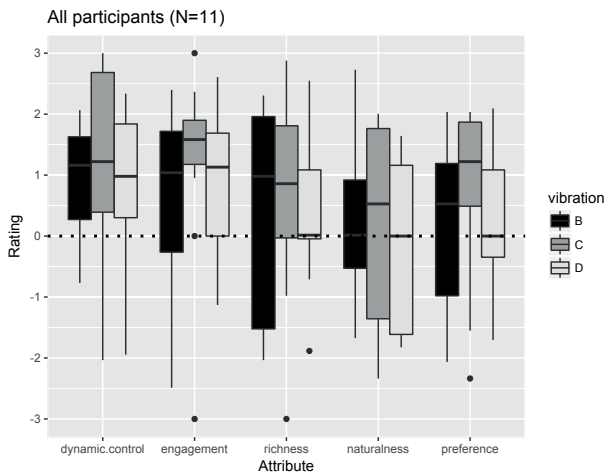


**Figura 4.** Confronto tra intensità delle vibrazioni (accelerazione RMS) registrate ai tasti del piano acustico suonati nell’esperimento e soglie vibrotattili (minime e massime) per tocco attivo misurate come descritto nella Sezione 2.1 (linee orizzontali).



**Figura 5.** La strumentazione utilizzata per la valutazione qualitativa di un piano digitale con feedback vibrotattile.

brazione realistica; C) vibrazione realistica incrementata di 9 dB; D) vibrazione sintetica. Più in dettaglio, la vibrazione realistica consisteva nei segnali di vibrazione registrati tramite un accelerometro da misura alla tastiera di un piano Disklavier a coda per 10 diversi livelli di MIDI velocity, mentre la vibrazione sintetica è stata prodotta al computer filtrando del rumore bianco con un filtro risonante impostato sulle frequenze fondamentali delle note corrispondenti, e intensità e involuppo di decadimento aderenti a quelli delle vibrazioni realistiche. Come negli esperimenti descritti nelle Sezioni 2.2 e 2.3, il feedback sonoro era fornito dal sintetizzatore di pianoforte Modart Pianoteq. L’ascolto è avvenuto attraverso cuffie isolanti, le quali prevenivano inoltre ai soggetti di percepire il suono inevitabilmente emesso dai trasduttori di vibrazione. Infine, sono state implementate alcune complesse procedure per calibrare la risposta dinamica, l’intensità e l’equalizzazione dei feed-



**Figura 6.** Valutazione della qualità del piano digitale in base alla presenza di vibrazioni. Valori positivi significano una preferenza per la configurazione con vibrazioni.

back sonoro e vibrotattile [6], così da rendere l’esperienza quanto più possibile simile a quella dell’esecuzione su pianoforte acustico.

Durante l’esperimento i pianisti potevano suonare liberamente, similmente a quanto descritto nella Sezione 2.2 per il piano acustico, commutando tra due configurazioni che comprendevano sempre la condizione A (assenza di vibrazione) e una delle condizioni con vibrazione (B, C o D). I soggetti dovevano quindi assegnare dei punteggi ai seguenti attributi legati allo strumento e all’esperienza della performance: controllo della dinamica, ricchezza sonora, coinvolgimento, naturalezza, preferenza generale.

Undici soggetti hanno partecipato all’esperimento, di età media pari a 26 anni e con un’esperienza media di 8 anni post-conservatorio.

#### 2.4.2 Risultati

I risultati, raccolti in Figura 6, mostrano che mediamente i soggetti hanno preferito le configurazioni con vibrazioni. L’attributo “naturalezza” nella condizione D (vibrazioni sintetiche) ha ricevuto numerosi punteggi negativi, come evidente dal rispettivo box-plot. Le preferenze più marcate sono state per gli attributi “controllo della dinamica” e “coinvolgimento”. La configurazione C (vibrazione realistica incrementata in intensità) è risultata in generale quella maggiormente preferita, con punteggi più elevati per 4 su 5 attributi. Curiosamente la configurazione B è risultata generalmente meno preferita di C suggerendo un effetto di assuefazione alla vibrazione simile a quanto avviene in ambito acustico per l’effetto “loudness”.

Analogamente all’esperimento di valutazione del piano acustico (Sezione 2.2), i soggetti hanno dimostrato una grande variabilità di opinione.

È interessante notare che, sebbene il feedback sonoro sia rimasto inalterato in tutte le configurazioni con e senza vibrazione, le preferenze per le prime sono state generalmente associate all’attributo “ricchezza sonora”. Ciò supporta l’ipotesi che la percezione della qualità di uno stru-

mento è un processo che integra sia informazione uditiva che aptica.

### 2.5 Influenza del feedback vibrotattile sulla performance con pianoforte digitale

Una delle questioni più rilevanti per l’aptica musicale è quella di verificare come e quanto il canale somatosensoriale supporti la manualità della pratica strumentale [17, 18, 19]. Per esempio in compiti musicali come l’esecuzione di un vibrato, l’intonazione, il raggiungimento di una dinamica prestabilita, e l’accuratezza temporale. Questa linea di ricerca presenta alcune delicate problematiche che finora sono state solo parzialmente esplorate: da un lato esiste una grande varietà di tipologie di strumenti musicali, ognuna delle quali offre possibilità di interazione uniche o peculiari, e dall’altro misurare in modo oggettivo l’esecuzione di un dato compito musicale rischia di escludere tutto ciò che riguarda l’espressività, snaturando il gesto musicale e riducendolo a qualcosa di puramente prestazionale.

L’ultimo esperimento qui riportato si situa appunto in questo contesto, e rappresenta un primo tentativo in questa direzione, lungi dall’essere esaustivo.

#### 2.5.1 Descrizione dell’esperimento

Questo esperimento aveva l’obiettivo di misurare alcuni parametri dell’esecuzione musicale sul pianoforte al variare del feedback vibrotattile. In particolare, si è misurata l’accuratezza temporale e dinamica nell’esecuzione di un semplice compito musicale.

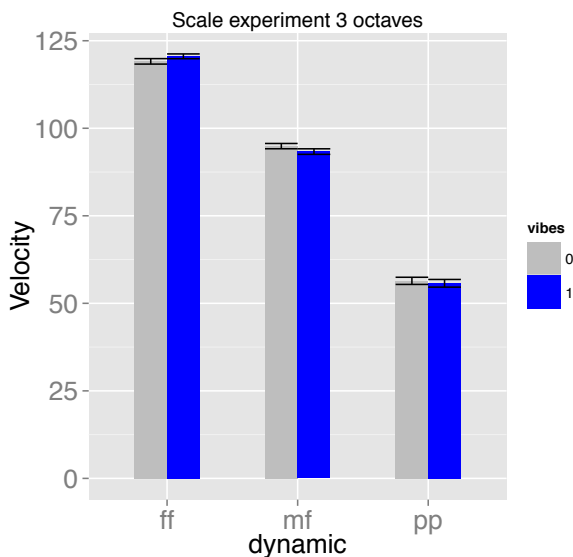
La strumentazione utilizzata era la stessa dell’esperimento descritto nella Sezione 2.4, e anche i partecipanti all’esperimento sono stati gli stessi che hanno preso parte alla valutazione qualitativa del piano digitale. Come condizioni sperimentali si sono considerate solo le configurazioni A e B, corrispondenti rispettivamente a vibrazione assente e vibrazione realistica. In aggiunta, i soggetti potevano udire il battito di un metronomo impostato a 120 BPM.

I soggetti dovevano eseguire una scala di re maggiore ascendente e poi discendente, seguendo il metronomo ad ogni secondo battito, e con una dinamica fissa assegnata tra *pp*, *mf*, e *ff*. Si sono considerate solo le prime tre ottave (esecuzione con la sola mano sinistra), dal momento che l’esperimento sulla percezione della vibrazione alla tastiera del pianoforte (Sezione 2.3) mostra che in quella regione la percezione è netta. I soggetti dovevano ripetere l’esecuzione 3 volte per ogni dinamica assegnata e in ogni condizione di feedback vibrotattile, per un totale di 18 ripetizioni in ordine casuale.

Il piano digitale era collegato ad un software realizzato in Pure Data che registrava i seguenti dati MIDI: “note on”, “note off” e “velocity”.

#### 2.5.2 Risultati

L’analisi dei dati MIDI ha permesso di verificare l’accuratezza temporale e di dinamica al variare delle condizioni sperimentali. Per quanto riguarda la misura di accuratezza temporale, si sono considerati gli intervalli tra successivi



**Figura 7.** Misure di accuratezza dinamica nell’esecuzione di una scala di RE maggiore sul piano digitale aumentato con e senza vibrazione.

eventi “note on” (*inter-onset intervals* o IOI) e la durata delle note.

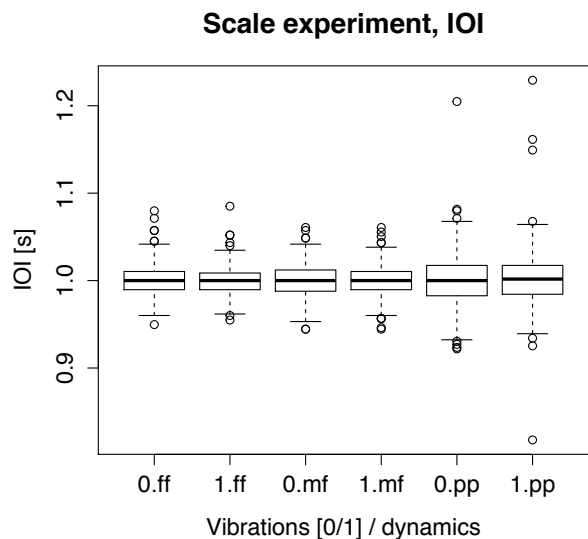
Come mostrato in Figura 7, per la dinamica assegnata *ff* si è verificato un leggero incremento nell’esecuzione con vibrazioni rispetto al caso senza vibrazioni, mentre è accaduto il contrario per la dinamica *mf*. Ad ogni modo, il contributo delle vibrazioni non è risultato statisticamente significativo né in generale né per le singole dinamiche separatamente.

La Figura 8 mostra che nemmeno l’accuratezza temporale data dalla grandezza IOI è stata influenzata significativamente dalla presenza di vibrazioni. La superiore variabilità registrata per la dinamica *pp* è probabilmente solo sintomo di maggiore difficoltà di esecuzione alle dinamiche basse. Anche la durata delle note è risultata stabile in tutte le condizioni vibrotattili.

In generale l’esperimento non ha portato risultati significativi. Una possibile motivazione sta nella facilità di esecuzione del compito assegnato, sia per la velocità del metronomo che per la semplicità di articolazione: effettivamente, qualsiasi pianista con un poco di esperienza può eseguire tale compito in modo del tutto automatico, probabilmente anche senza feedback uditivo o visivo. L’autore sta attualmente lavorando ad un nuovo esperimento analogo, che permetta di isolare in modo più accurato l’eventuale apporto dato dal feedback vibrotattile nell’esecuzione di compiti musicali.

### 3. CONCLUSIONI

La comprensione di come il canale sensoriale aptico supporti la pratica strumentale e contribuisca alla valutazione degli strumenti musicali richiede ancora un sostanziale apporto da parte della ricerca in ambiti sia teorici che applicati, quali la fisiologia e la psicofisica della somato-



**Figura 8.** Misure di accuratezza temporale nell’esecuzione di una scala di RE maggiore sul piano digitale aumentato con e senza vibrazione.

sensazione, la tecnologia aptica, il design e la valutazione dell’interazione multimodale uomo-macchina.

Nonostante l’aptica musicale sia per molti versi solo agli inizi, sempre più risultati sperimentali, anche se parziali, confermano che il canale aptico ricopre un ruolo di prima importanza nello scenario musicale. Lo sviluppo di futuri strumenti e interfacce musicali digitali che forniscano feedback aptico avanzato è quindi una priorità per supportare la manualità e l’espressività della performance con tali dispositivi, così come per migliorare la loro qualità percepita e l’esperienza generale del musicista.

### 4. RINGRAZIAMENTI

Si ringraziano le persone che hanno collaborato a vario titolo alle ricerche qui descritte, in particolar modo (in ordine alfabetico): Prof. Federico Avanzini, Prof. Federico Fontana, Martin Fröhlich, Hanna Järveläinen, Sebastien Schiesser.

Queste ricerche sono state parzialmente finanziate dal Fondo Nazionale Svizzero per la ricerca scientifica tramite i progetti 150107 “Audio-Haptic modalities in Musical Interfaces” e 178972 “Haptic technology and evaluation for digital musical interfaces”.

### 5. REFERENCES

[1] A. Del Piccolo, D. Rocchesso, and S. Papetti, “Path following in non-visual conditions,” *IEEE Trans. on Haptics*, 2018.

[2] S. Papetti and C. Saitis, “Musical haptics: Introduction,” in *Musical Haptics* (S. Papetti and C. Saitis, eds.), pp. 1–7, Cham: Springer Int. Publishing, 2018.

- [3] G. Békésy, “Über die vibrationsempfindung [on the sensation of vibration],” *Akustische Zeitschrift*, pp. 316–334, 1939.
- [4] S. J. Bolanowski, G. A. Gescheider, R. T. Verrillo, and C. M. Checkosky, “Four channels mediate the mechanical aspects of touch,” *J. Acoust. Soc. Am.*, vol. 84, pp. 1680–94, Nov. 1988.
- [5] R. T. Verrillo, “Vibration sensation in humans,” *Music Percept.*, vol. 9, no. 3, pp. 281–302, 1992.
- [6] S. Papetti, M. Fröhlich, F. Fontana, S. Schiesser, and F. Avanzini, “Implementation and characterization of vibrotactile interfaces,” in *Musical Haptics* (S. Papetti and C. Saitis, eds.), pp. 257–282, Cham: Springer Int. Publishing, 2018.
- [7] H. Levitt, “Transformed Up-Down methods in Psychoacoustics,” *J. Acoust. Soc. Am.*, vol. 49, no. 2, pp. 467–477, 1971.
- [8] S. O’Modhrain and R. B. Gillespie, “Once more, with feeling: Revisiting the role of touch in performer-instrument interaction,” in *Musical Haptics* (S. Papetti and C. Saitis, eds.), pp. 11–27, Cham: Springer Int. Publishing, 2018.
- [9] C. Saitis, H. Järveläinen, and C. Fritz, “The role of haptic cues in musical instrument quality perception,” in *Musical Haptics* (S. Papetti and C. Saitis, eds.), pp. 73–93, Cham: Springer Int. Publishing, 2018.
- [10] F. Fontana, S. Papetti, H. Järveläinen, and F. Avanzini, “Detection of keyboard vibrations and effects on perceived piano quality,” *J. Acoust. Soc. Am.*, vol. 142, no. 5, pp. 2953–67, 2017.
- [11] H. Suzuki, “Vibration and sound radiation of a piano soundboard,” *J. Acoust. Soc. Am.*, vol. 80, no. 6, pp. 1573–82, 1986.
- [12] M. Keane and G. Dodd, “Subjective Assessment of Upright Piano Key Vibrations,” *Acta Acust. united with Acust.*, vol. 97, pp. 708–713, July 2011.
- [13] A. Askenfelt and E. V. Jansson, “On vibration and finger touch in stringed instrument playing,” *Music Perception*, vol. 9, no. 3, pp. 311–350, 1992.
- [14] F. Fontana, F. Avanzini, H. Järveläinen, S. Papetti, F. Zanini, and V. Zanini, “Perception of Interactive Vibrotactile Cues on the Acoustic Grand and Upright Piano,” in *Proc. Sound Music Comput. Conf. (SMC)*, (Athens, Greece), 2014.
- [15] D. Green and J. Swets, *Signal detection theory and psychophysics*. New York: Wiley and Sons, 1966.
- [16] F. Fontana, F. Avanzini, H. Järveläinen, S. Papetti, G. Klauer, and L. Malavolta, “Rendering and subjective evaluation of real vs. synthetic vibrotactile cues on a digital piano keyboard,” in *Proc. Sound Music Computing conf. (SMC)*, (Maynooth, Ireland), 2015.
- [17] W. Goebel and C. Palmer, “Tactile feedback and timing accuracy in piano performance,” *Exp. Brain Res.*, vol. 186, pp. 471–479, 2008.
- [18] T. Ahmaniemi, “Effect of dynamic vibrotactile feedback on the control of isometric finger force,” *IEEE Trans. on Haptics*, 2012.
- [19] H. Järveläinen, S. Papetti, S. Schiesser, and T. Groschauser, “Audio-tactile feedback in musical gesture primitives: Finger pressing,” in *Proceedings of the Sound and Music Computing Conference (SMC2013)*, (Stockholm), 2013.

# Dynamic Preservation of Interactive Art: The next frontier of Multimedia Cultural Heritage (DaphNet)

Federica Bressan

IPEM - Institute for Psychoacoustics and Electronic Music  
Dept. of Art history, Music and Theatre studies  
Ghent University, Belgium  
federica.bressan@ugent.be

## 1. INTRODUCTION

Interaction with multimedia technology for artistic activities is a fairly young but intriguing domain, holding the promise of a fundamental innovation of the European art scene. Its development is linked with innovative trends in the academic approach to digital humanities, and this has strong links with the industrial sectors of virtual reality, gaming, medical rehabilitation and learning. However, despite the fact that Europe holds the biggest cultural patrimony in the world, and that its artists are very innovative, it is the United States of America that currently lead the research on art preservation (including archival management and related digital technologies). The academic curricula on preservation offered in the US greatly out-number those offered in Europe. As a result, overseas, professional figures with a multidisciplinary profile between the arts and computer science are recognised, competitive and in demand.

Europe has the potential to re-gain a leading role by exploiting the intrinsic value of its culture and creativity, which “enables development and empowers people, with non-monetised social benefit”<sup>1</sup>. The DaphNet project I aspire to show that high-profile multidisciplinary professional figures have the power to accelerate and increase the economic value of the European cultural heritage. The sector in which these figures operate cuts across the arts, business and connectivity, driving innovation and new business models, with immediate spillovers in adjacent sectors such as artistic production, scientific dissemination, education and museum design. In addition, I want to prove that a young woman can achieve this professional status.

These goals meet the priorities of the Work Programme H2020 on the canalisation of advanced competences for the safeguard and the promotion of European cultural heritage, and on interdisciplinary innovation for the European cultural and creative sectors.

<sup>1</sup> UNESCO, Creative Economy Report 2013 Special Edition, 2013, Paris (France), p. 9, ISBN 978-92-3-001211-3

Copyright: ©2010 Federica Bressan et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution License 3.0 Unported](https://creativecommons.org/licenses/by/3.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

## 2. STATE OF THE ART

A major problem with the preservation of interactive art, moreover, is that established archival practices require that objects are made static for preservation, failing to capture the essence of an art form that is dynamic and transitory. During the past 10 years the archival community has grown awareness about the specific problems of interactive art preservation, starting dedicated workshops and symposia<sup>2</sup>, and some research projects<sup>3</sup> have shaped extremely interesting concepts. However, the challenge is to produce practical and re-usable shared output for the description of interaction with artistic installations, so that installations can be re-build even if the technologies have become obsolete. The practice of interactive art preservation today remains unsatisfactory, open to ad hoc fixes that may save the installation for the next exhibition, but certainly not for the next generations to see, experience and learn. An effective model for the preservation of interactive art is currently missing. Yet research shows that the intrinsic use of computer-based information technologies in art projects requires a multidisciplinary approach to preservation. DaphNet aims at contributing to a breakthrough solution for this problem, by investigating the processes on which human interaction with multimedia interactive art are based.

The main objective of this project is to define a framework that describes how people interact with multimedia art installations. This framework will comprise an ontology for interactive art, and it will be based on empirical work on artistic interaction. The goal is then to use this in the model for interactive art preservation, so that a re-build of the artwork, even with other technologies, can keep the essence of the artistic interaction.

<sup>2</sup> For example: “Digital Oblivion” organised in 2010 and 2011 by the ZKM (Zentrum für Kunst und Medientechnologie) in Karlsruhe, Germany.

<sup>3</sup> [1] The Variable Media Initiative at the Guggenheim Foundation (1999); [2] R. Fromm et al., Capturing Unstable Media Arts. A formal model for describing and preserving aspects of electronic media art, chapter in “Present Continuous Past(s)”, Springer Vienna, 2005, pp 174-189; [3] Inside Installations. Theory And Practice In The Care Of Complex Artworks, T. Scholte and G. Wharton (eds.), Amsterdam Univ. Press, 2011; [4] Matters in Media Art (since 2005), with the New Art Trust (NAT) and its partner museums ? Museum of Modern Art (MoMA), San Francisco Museum of Modern Art (SFMOMA) and Tate Museum in London (<http://www.tate.org.uk/about/projects/matters-media-art>); [5] Experience in the Archive. The central role of human experience [with] new media art (2007) with Lizzie Muller at the French Canadian private Foundation Daniel Langlois (<http://www.fondation-langlois.org/html/e/page.php?NumPage=2096>).

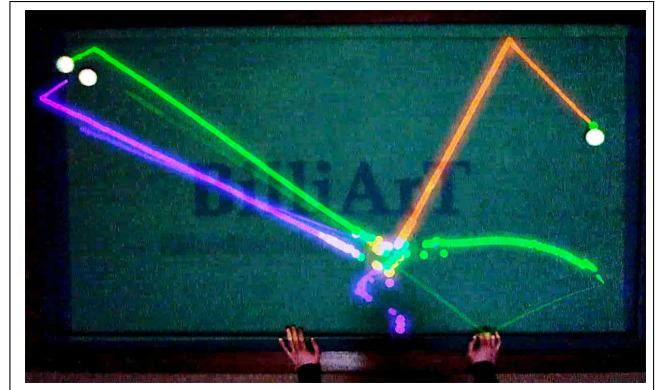
DaphNet is interdisciplinary because it integrates knowledge and instruments from different disciplines to create a new model that could not occur if these were handled separately. The added value of multidisciplinary resides in the ability of switching contexts, and the competences represented in DaphNet allow for the connection between the worlds of the humanities with engineering, with immediate societal benefits by preserving artwork that might otherwise be lost, producing more cost-effective and sustainable models that address current archiving needs, and creating business opportunities for new technologies and services.

DaphNet is relevant because it addresses an important new development in the European culture, which we are currently failing to preserve and promote/capitalise. Due to its influence on other art forms as well as on the contemporary culture, the evolution of interactive installation art may be considered as important as the evolution of opera in the XVII century. The work during DaphNet would unlock the potential that interactive art has to inform and transform other cultural and industrial sectors, attracting more young researchers to engage in this field, and opening new career opportunities for young professionals.

DaphNet is timely due to the fast decay of multimedia artworks (with a life expectancy often limited to the time of the exhibition), and to the heavy reliance of these artworks on computer-based technology – making it subject to rapid obsolescence. Expertise in digital archaeology is required to recuperate artworks produced since the 1990s, as media (e.g. floppy disks), programming languages, hardware and software have already been replaced by several generations of technology.

### 3. METHODOLOGY AND APPROACH

Notable recent studies carried out at IPPEM are the basis for the theoretical ontology and taxonomy of artworks, reflecting the type and degree of interaction designed by the artist. I will develop a structured data storage system to ensure optimised management during the project, maximised potential for dissemination, and persistence of the data after the project. Experimental pilots will then be carried out in ecological settings, involving artworks provided by the members of the network and users. I will extensively interview artists and other actors involved in the world of the arts, producing curated multimedia documentation. The subjective data will be coupled with objective measurements, including biometric sensors, motion sensors and other sensing technologies. If necessary, new artworks with specific features functional to the present study will be commissioned to in-house artists at the IPPEM centre, who already agreed to collaborate. The above methods are all standard at IPPEM.



**Figure 1.** BilliArT installation at IPPEM. Balls trajectories graphically reconstructed with the motion capture data log, and superimposed to the real-life video captured from the top of the table.

The originality of this research work resides in: (1) the topic of the investigation, i.e. multi-modal interaction in the specific context of artistic installations; (2) in the multidisciplinary approach combining the arts, computer science and digital humanities; and finally, (3) in the integration of subjective and objective data derived from empirical observation. The main innovative aspects of my research are (i) the application of the embodied cognition paradigm to this research topic, and (ii) the exploitation of the results of this project in the field of cultural heritage preservation.

#### 3.1 Gender mainstreaming

According to the definition of gender mainstreaming<sup>4</sup>, gender analysis is cross-integrated at macro, meso and micro levels throughout this project cycle. An extensive study on the representation of women in the world of the arts (as artists, curators, promoters, etc.) is currently missing. By carrying out my research on interactive installations, I intend to address the problem by: (1) collecting quantitative data and statistics from the cultural institutions involved in my project and the members of the project network; (2) monitoring significant indicators with focused questions in interviews, life stories and questionnaires. In addition, (3) I will make sure that at least 40% of the participants involved in the experiments are women; and (4) in the selection of artworks, I will assign a priority to women's works? granted that the characteristics of the installations meet the requirements of the study. In all these activities, I will be aided by experts in the field, whom I have already contacted personally and whom have agreed to give their contribution.

<sup>4</sup> EU Commission. Manual for gender mainstreaming. Employment, social inclusion and social protection policies, 2008.

# A Holistic Approach to the Design of Digital Musical Instruments

Fabio Morreale

Augmented Instruments Lab

Centre for Digital Music

Queen Mary University of London

f.morreale@qmul.ac.uk

## ABSTRACT

The number of new musical instruments has been rapidly growing in the last few years. Yet, few have achieved lasting impact beyond the first few performances. A few attempts have been made to solve this issue; most of them addressing it from a design. I argue that framing this issue only from a design perspective might fail to address the complexity of the problem at hand. Theory and praxis of musical interface design would benefit from a holistic approach, combining knowledge and methods from different disciplines, among which HCI, Music Cognition, and Philosophy.

## 1. INTRODUCTION

Many new digital musical instruments (DMIs) are invented every year, but nearly all of them fail to go mainstream [3]. This issue has been mostly addressed from a design perspective, leading to the proposition of numerous design frameworks for DMIs to offer instrument designers a theoretical grounding [2,5].

I argue that framing this issue only from a design perspective might fail to address the complexity of the problem at hand. Theory and praxis of musical interface design, I contend, would benefit from a holistic approach, combining knowledge and methods from different disciplines - among which Human-Computer Interaction (HCI), Music Cognition, and Philosophy. I will make this case by presenting a number of projects that we have been working on at the Augmented Instruments Lab of Queen Mary University of London.

From a HCI perspective, we explored design and performance practices of DMI designers and performers and proposed design practices that should be followed to facilitate a prolonged use of musical instruments [3,4]. This study suggested that a substantial problem causing a limited uptake of new instruments is musician's need to re-learn from scratch a new instrument.

This issue motivated us to investigate how to repurpose the existing expertise of trained musicians in the design of new DMIs, so that new creative possibilities are

*Copyright: © 2018 First author's last name et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution License 3.0 Unported](https://creativecommons.org/licenses/by/3.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.*

offered without requiring thousands of hours of retraining. We addressed this issue from a sensorimotor skills acquisition perspective. At expert level, the instrument becomes *transparent* to the performer in a way that the instrumental operations become automatic. We performed an experimental study with expert violinists aimed to identify the extent to which the design of the instrument can be altered before this transparency effect fades away.

Dealing with instrument transparency allowed us to generalise this issue and look into the experience of humans (musicians) embodying technologies (instruments). This connection opened up links with other disciplines. In particular, we encountered and got fascinated by theories and methodologies of postphenomenology [1], a philosophy of technology that studies with how humans interact with technology. For our latest study, we then performed a postphenomenological investigation to study how musicians interacted with a newly developed instrument [6]. We created an augmented guitar pick and had expert guitarist to use it for a period of time. We were interested in understanding how the specificities of our technology shaped the practice of the guitarists and, in turn, investigating the implications of such practices to offer new bases for knowledge development.

## 2. REFERENCES

- [1] D. Ihde. "Technology and the lifeworld: From garden to earth". Indiana University Press, 1990.
- [2] F. Morreale, A. De Angeli, and S. O'Modhrain. "Musical interface design: An experience-oriented framework". In *Proceedings of New Interfaces for Musical Expression*, 2014.
- [3] F. Morreale and A. McPherson. "Design for longevity: Ongoing use of instruments from NIME 2010-14". In *Proceedings of New Interfaces for Musical Expression*, 2017.
- [4] F. Morreale, A. McPherson, and M. Wanderley. "NIME Identity from the Performer's Perspective". In *Proceedings of New Interfaces for Musical Expression*, 2018.
- [5] D. Overholt. "The musical interface technology design space". *Organised Sound*, 14(2), 2009.



# UN MODELLO FISICO PER PIANOFORTE PREPARATO

**Michele Ducceschi**

Acoustics and Audio Group

University of Edinburgh

Scotland, UK

michele.ducceschi@ed.ac.uk

## ABSTRACT

Questo lavoro presenta uno schema numerico per la simulazione di un pianoforte preparato. Tale strumento musicale si compone di un normale pianoforte, a cui possono essere applicati a piacere degli strumenti di preparazione. Tali elementi (sonagli, ammortizzatori, chiodi), supposti puntiformi, entrano in contatto per brevi istanti con la corda, arricchendone il contenuto armonico. Vista la natura estremamente non lineare del sistema, uno schema numerico implicito verrà proposto, la cui stabilità verrà inferita dalla conservazione di un'energia discreta non negativa, assimilabile all'hamiltoniano del sistema.

## 1. INTRODUZIONE

Il pianoforte preparato ha una storia affascinante e ricca. Uno dei primi esempi di tale strumento si deve a Delage, il quale collocò del cartone (un ammortizzatore) al di sotto di alcune corde per ottenere dei suoni da percussione indiana. Un altro famoso esempio si deve a Cowell, il quale era solito operare direttamente sulle corde e sulla tavola, con vari oggetti. Probabilmente, gli esempi più noti di preparazione sono dovuti a John Cage, autore di preparazioni eccellenti facenti uso degli strumenti più vari tra i quali figurano chiodi, monetine e posate [1].

In questo lavoro, un modello completo di pianoforte verrà presentato. Esso è composto dalle corde, dalla tavola armonica alla quale le corde trasmettono energia attraverso il ponticello, dai martelletti, e dagli strumenti di preparazione. Nella sezione 2 verranno presentate le equazioni continue per tutti gli elementi, e le loro interconnessioni. Nella sezione 3 le controparti discrete saranno elaborate. Lo schema numerico si compone di una parte a differenza finita (per la corda, i martelletti e gli strumenti di preparazione), e di una parte modale (per la tavola armonica). L'accoppiamento dei due sistemi numerici avviene attraverso il ponticello, in maniera esplicita. In modo da garantire stabilità, lo schema numerico sarà tale da conservare un hamiltoniano discreto non negativo. Infine, la sezione 4 conterrà degli esperimenti numerici. Esempi sonori

associati a questo lavoro possono essere ascoltati sul sito personale dell'autore <sup>1</sup>.

## 2. MODELLO CONTINUO

In questa sezione verranno presentati i modelli continui delle componenti del pianoforte preparato. Il pianoforte è composto da risonatori primari, le corde, ovvero sistemi unidimensionali descritti da un'opportuna equazione delle onde. Tali risonatori sono messi in vibrazione da oggetti puntiformi, i martelletti, attraverso una collisione che avviene in un breve arco di tempo. La vibrazione delle corde è trasmessa alla tavola armonica attraverso un ponticello. La tavola armonica agisce da risonatore secondario. Punti di ascolto virtuali sono collocati sulla tavola armonica, oltre che sulle corde nelle direzioni trasversale e longitudinale. Per la preparazione del pianoforte, oggetti tipo sonagli, chiodi e ammortizzatori possono essere collocati a piacere sulla tavola, in modo da entrare in contatto con le corde a seconda dell'ampiezza delle vibrazioni di queste ultime. Chiaramente, la natura non lineare delle collisioni e dei contatti darà luogo a vibrazioni dal ricco contenuto spettrale, che andranno a modificare il colore tipico del suono del pianoforte.

### 2.1 Le corde

Le corde del pianoforte agiscono come risonatori primari dello strumento. Esse infatti vengono messe in moto dal martelletto, e trasferiscono l'energia di vibrazione alla tavola armonica attraverso il ponticello. I modelli fisici di corda vibrante sono numerosi in letteratura, a partire dal più semplice modello descritto dalla equazione delle onde unidimensionale [2], fino a modelli di Timoshenko nonlineari geometricamente esatti [3]. I due esempi citati rappresentano dei casi limite: il primo è troppo semplice per rendere conto dei fenomeni fisici fondamentali per una sintesi realistica, l'altro è fin troppo complesso (si veda, a questo proposito, [4]). Per il pianoforte, è noto che il contributo alla forza al ponticello è dovuto sia alle vibrazioni trasversali delle corde, sia a quelle longitudinali [5]. L'inarmonicità delle vibrazioni trasversali è importante, perché dà origine al fenomeno delle "ottave allungate" (stretched octaves in inglese) tipico del piano, come è importante un certo livello di non linearità intrinseca per accoppiare le vibrazioni trasversali a quelle longitudinali. Un modello

Copyright: ©2018 Michele Ducceschi et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution License 3.0 Unported](https://creativecommons.org/licenses/by/3.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

<sup>1</sup> All'indirizzo [mdphys.org](http://mdphys.org), nella sezione Prepared Piano.

soddisfacente, in grado di descrivere tutti i fenomeni appena citati, è il modello non lineare di Morse e Ingard [6]. In assenza di dissipazione, esso è descritto dalla seguente coppia di equazioni differenziali alle derivate parziali

$$\mathcal{L}(w) = \mathcal{N}\mathcal{L}(w, \zeta) + \sum_i g_i \mathcal{F}_i \quad (1a)$$

$$\mathcal{G}(\zeta) = \mathcal{N}\mathcal{G}(w) \quad (1b)$$

Nelle equazioni,  $w(x, t)$  e  $\zeta(x, t)$  rappresentano, rispettivamente, le componenti trasversale e longitudinale del moto della corda. Gli operatori  $\mathcal{L}, \mathcal{G}$  sono lineari, e dati da

$$\mathcal{L}(w) = \rho_c A \partial_t^2 w - T_0 \partial_x^2 w + EI \partial_x^4 w \quad (2a)$$

$$\mathcal{G}(\zeta) = \rho_c A \partial_t^2 \zeta - EA \partial_x^2 \zeta \quad (2b)$$

Gli operatori  $\mathcal{N}\mathcal{L}, \mathcal{N}\mathcal{G}$  sono invece non lineari, e dati da

$$\mathcal{N}\mathcal{L}(w, \zeta) = \frac{EA - T_0}{2} \partial_x [(\partial_x w)^3 + 2\partial_x w \partial_x \zeta] \quad (3a)$$

$$\mathcal{N}\mathcal{G}(w) = \frac{EA - T_0}{2} \partial_x [(\partial_x w)^2] \quad (3b)$$

Il simbolo  $\partial_s^n$  rappresenta una derivata parziale di ordine  $n$  secondo la variabile  $s$ . La tensione applicata alla corda è denominata  $T_0$  nelle equazioni; inoltre  $E$  è il modulo di Young e  $I$  il momento di inerzia della sezione di taglio (supposta circolare); infine,  $\rho_c$  è la densità volumetrica (ove il pedice  $c$  indica corda, per distinguerla dalla densità della tavola armonica), e  $A$  l'area della sezione di taglio. Infine,  $g_i$  rappresenta una distribuzione temperata che quantifica l'estensione spaziale della forza  $F_i$  agente sulla corda. I vari tipi di forze verranno specificati nelle sezioni seguenti, e si intendono provenienti dai contatti con il martelletto e con gli strumenti di preparazione.

Se si assume regolarità delle condizioni iniziali, in assenza dei termini di forza, è possibile dimostrare l'esistenza di una soluzione forte per il sistema (23). Perciò,  $w, \zeta$  sono sufficientemente regolari e per esse è possibile definire un prodotto scalare e norma associata del tipo

$$\langle a, b \rangle_{\mathcal{D}} \triangleq \int_0^L ab \, dx \quad (4a)$$

$$\|a\|_{\mathcal{D}}^2 \triangleq \int_0^L a^2 \, dx \quad (4b)$$

ove la corda è definita nell'intervallo chiuso  $\mathcal{D} \triangleq [0, L]$ .

Le espressioni delle forze trasversale ( $\mathcal{F}_t$ ) e longitudinale ( $\mathcal{F}_l$ ) che sono trasmesse alla tavola armonica attraverso il ponticello, hanno le seguenti espressioni

$$\mathcal{F}_t = \mathcal{F}_t^{(l)} + \mathcal{F}_t^{(nl)} \quad (5a)$$

$$\mathcal{F}_l = \mathcal{F}_l^{(l)} + \mathcal{F}_l^{(nl)} \quad (5b)$$

$$\mathcal{F}_t^{(l)} = T_0 \partial_x w - EI \partial_x^3 w \quad (5c)$$

$$\mathcal{F}_t^{(nl)} = \frac{EA - T_0}{2} [(\partial_x w)^3 + 2\partial_x w \partial_x \zeta] \quad (5d)$$

$$\mathcal{F}_l^{(l)} = EA \partial_x \zeta \quad (5e)$$

$$\mathcal{F}_l^{(nl)} = \frac{EA - T_0}{2} (\partial_x w)^2 \quad (5f)$$

Le espressioni di queste forze discendono dalla conservazione dell'energia del sistema, come sarà dimostrato in 2.6. Come si può notare, le forze contengono termini non lineari. Essi sono importanti in quanto danno origine al fenomeno delle "armoniche fantasma" (phantom partials in inglese), ovvero di armoniche non presenti nello spettro della corda che vibra in regime lineare. In lavori precedenti, solo l'accoppiamento da trasversale a longitudinale è stato considerato, con risultati molto soddisfacenti dal punto di vista della sintesi [7]. Qui, il modello verrà mantenuto completo in modo da permettere di suonare lo strumento in regime fortemente non lineare, qualora lo si desidererà.

## 2.2 Il martelletto

Il martelletto del pianoforte gioca un ruolo fondamentale nella produzione del suono. Esso è costituito da una struttura in legno ricoperta da diversi strati di feltro di lana. La durezza dei martelletti cambia in funzione della corda su cui agiscono, ed in genere le corde più acute vengono messe in vibrazione da martelletti di maggiore durezza. A livello dinamico, è possibile modellare l'azione del martelletto come una molla non lineare, con una legge del tipo

$$\mathcal{F}_m = K_m [\eta_m]_+^{p_m}, \quad (6)$$

ove  $\eta_m$  rappresenta la compressione del martelletto in contatto con la corda,  $p_m \geq 1$ , e  $K_m$  è una costante elastica. Il simbolo  $[\cdot]_+$  indica la parte positiva, cioè

$$[x]_+ = \frac{x + |x|}{2} \quad (7)$$

Tale modello non rende conto del fenomeno di isteresi che si osserva nel contatto tra corda e martelletto [8], ma rimane nonostante un modello realistico per la sintesi sonora, utilizzato in molti lavori precedenti [3, 7, 9]. La variabile  $\eta_m$  può essere espressa in termini della posizione del martelletto, e della corda, come nella seguente

$$\eta_m = U_m - \langle g_m, w \rangle_{\mathcal{D}}, \quad (8)$$

ove  $U_m$  è la posizione del martelletto (supposto agente al di sotto della corda), e  $g_m$  è una distribuzione temperata tale che  $\langle g_m, 1 \rangle_{\mathcal{D}} = 1$  (ad esempio, una delta di Dirac). La forza è conservativa, e descritta dal potenziale

$$\phi(\eta_m) = \frac{K_m}{p_m + 1} [\eta_m]_+^{p_m + 1} \quad (9)$$

L'equazione del moto del martelletto è dunque

$$M_m \partial_t^2 U_m = -\mathcal{F}_m \quad (10)$$

## 2.3 La tavola armonica

La tavola armonica del pianoforte agisce come risonatore secondario. Nei pianoforti a coda, essa è di forma complessa, incastrata dentro ad una cornice rigida. Il materiale di cui è costituita è solitamente legno, che è un materiale ortotropico. Delle coste in legno vengono applicate alla tavola per cambiarne la rigidità locale, e per conferire una maggiore resistenza agli sforzi usuranti delle corde. Evidentemente la modellazione di una tavola armonica di tale

complessità richiede uno sforzo matematico non banale, oltre che avanzate tecniche di risoluzione numerica [10]. È tuttavia possibile creare un modello calibrato su risposte all'impulso provenienti da misure sperimentali [7]. Per lo scopo di questo lavoro, è conveniente cercare di modellare la tavola in maniera semplice, ma sufficientemente realistica, in modo da poter operare in libertà su alcuni parametri fisici di rilievo, ad esempio i tempi di decadimento. Un modello sufficiente allo scopo è il modello di Kirchhoff di piastra sottile ortotropica che, in assenza di dissipazione e di forze esterne, si scrive

$$\mathcal{T}(W) = -g_p^{(t)} \mathcal{F}_t - g_p^{(l)} \mathcal{F}_l \quad (11a)$$

$$\mathcal{T}(W) = \rho_p H \partial_t^2 W - T_p [\partial_x^2 + \partial_y^2] W + \partial \mathcal{D} W \quad (11b)$$

$$\partial \mathcal{D} = [D_x \partial_x^4 + D_y \partial_y^4 + 2D_{xy} \partial_x^2 \partial_y^2] \quad (11c)$$

dove  $\rho_p$  è la densità volumetrica della piastra,  $H$  è lo spessore,  $T_p$  la tensione applicata,  $D_x, D_y, D_{xy}$  sono le costanti di rigidità. Esse hanno le seguenti espressioni

$$D_x = \frac{E_x H^3}{12(1 - \nu_x \nu_y)} \quad (12a)$$

$$D_y = \frac{E_y H^3}{12(1 - \nu_x \nu_y)} \quad (12b)$$

$$D_{xy} = D_x \nu_y + \frac{GH^3}{6} \quad (12c)$$

dove  $E_x, E_y$  sono i moduli di Young secondo le direzioni ortogonali  $x, y$ .  $\nu_x, \nu_y$  sono le costanti di Poisson, e  $G$  è il modulo di rigidità. Anche per  $W$ , soluzione di (11a), è possibile definire un prodotto scalare e una norma associata, del tipo

$$\langle a, b \rangle_{\mathcal{A}} \triangleq \int_0^{L_x} \int_0^{L_y} ab \, dx dy \quad (13a)$$

$$\|a\|_{\mathcal{A}}^2 \triangleq \int_0^{L_x} \int_0^{L_y} a^2 \, dx dy \quad (13b)$$

dove  $\mathcal{A} \triangleq [0, L_x] \times [0, L_y]$ . In (11a),  $g_p^{(t)}, g_p^{(l)}$  sono delle distribuzioni temperate che, come nel caso del martelletto in (8), soddisfano  $\langle g_p^{(t)}, 1 \rangle_{\mathcal{A}} = \langle g_p^{(l)}, 1 \rangle_{\mathcal{A}} = 1$ . Le espressioni delle forze  $\mathcal{F}_t, \mathcal{F}_l$  sono date da (5).

## 2.4 Il ponticello

Il ponticello ha lo scopo di trasmettere le vibrazioni della corda alla tavola armonica. Il ponticello può essere modellato come un sistema multimodale la cui risposta in frequenza può essere dedotta da misure sperimentali. Un modello estremamente semplificato verrà utilizzato in questo lavoro: la corda è connessa alla tavola attraverso due molle, per le componenti longitudinale e trasversale del moto. Riprendendo le espressioni delle forze agenti sul capo della corda, (5), si pone

$$\mathcal{F}_t = -K_t \left( w(L) - \langle g_p^{(t)}, W \rangle_{\mathcal{A}} \right) \quad (14a)$$

$$\mathcal{F}_l = -K_l \left( \zeta(L) - \langle g_p^{(l)}, W \rangle_{\mathcal{A}} \right) \quad (14b)$$

dove  $K_t, K_l$  sono le costanti elastiche.

## 2.5 Gli strumenti di preparazione

La preparazione del pianoforte avverrà posizionando elementi non lineari lungo la corda, con estensione spaziale descritta da opportune distribuzioni temperate  $g_i$  come in (23a). Tali elementi (che potranno essere assimilati a chiodi, sonagli, e ammortizzatori) imprimono alla corda una forza al momento del contatto. Per i chiodi (posti al di sotto della corda, e supposti immobili) vale la seguente

$$\mathcal{F}_c = K_c [\eta_c]_+^{p_c}, \quad \eta_c = U_c - \langle g_c, w \rangle_{\mathcal{D}} \quad (15)$$

Per gli ammortizzatori, essi soddisfano

$$\mathcal{F}_a = -\beta \partial_t \eta_a [\eta_a]_+^{p_a}, \quad \eta_a = U_a - \langle g_a, w \rangle_{\mathcal{D}} \quad (16)$$

dove  $\beta > 0$ . I sonagli sono elementi mobili, che agiscono sia sopra che sotto la corda, e descritti da

$$\mathcal{F}_s = \begin{cases} -K_s (\eta_s - b_s)_+^{p_s} & \text{if } \eta_s > b_s \\ 0 & \text{if } |\eta_s| \leq b_s \\ K_s (-\eta_s - b_s)_+^{p_s} & \text{if } \eta_s < -b_s \end{cases} \quad (17)$$

dove  $\eta_s = \langle g_s, w \rangle_{\mathcal{D}} - U_s$  e la cui equazione del moto soddisfa

$$M_s \partial_t^2 U_s = -\mathcal{F}_s. \quad (18)$$

## 2.6 Conservazione dell'energia del sistema continuo

Il sistema completo appena descritto è conservativo. Come esempio, per semplicità, verrà adesso considerato uno strumento composto da una sola corda e senza strumenti di preparazione. La generalizzazione a un sistema multicorda, e preparato, risulterà immediata. Si consideri dunque una corda, un martelletto, e la tavola armonica. Il sistema si scrive

$$\mathcal{L}(w) = \mathcal{N} \mathcal{L}(w, \zeta) + g_m \mathcal{F}_m \quad (19a)$$

$$\mathcal{G}(\zeta) = \mathcal{N} \mathcal{G}(w) \quad (19b)$$

$$\mathcal{T}(W) = -g_p^{(t)} \mathcal{F}_t - g_p^{(l)} \mathcal{F}_l \quad (19c)$$

$$M_m \partial_t^2 U_m = -\mathcal{F}_m \quad (19d)$$

Il sistema (19) è Hamiltoniano. Per il sistema, si considerino distribuzioni temperate tipo delta di Dirac per  $g_m, g_p^{(t)}, g_p^{(l)}$ . Si prendano i prodotti scalari, come definiti in (4), di (19a) con  $\partial_t w$ , e di (19b) con  $\partial_t \zeta$ . Si prenda un prodotto scalare, come definito in (13), di (19c) con  $\partial_t W$ . Infine si moltiplichi (19d) per  $\partial_t U_m$ . Dopo opportune integrazioni per parti, e dopo aver sommato i contributi delle equazioni, si ottiene il seguente bilancio energetico

$$\frac{d}{dt} \mathcal{H} = 0 \quad (20)$$

dove

$$\begin{aligned}
 \mathcal{H} &= \mathcal{H}^{(c)} + \mathcal{H}^{(p)} + \mathcal{H}^{(m)} + \mathcal{H}^{(b)} \\
 \mathcal{H}^{(c)} &= \mathcal{E}_k^{(c)} + \mathcal{E}_{pl}^{(c)} + \mathcal{E}_{pnl}^{(c)} \\
 \mathcal{H}^{(p)} &= \mathcal{E}_k^{(p)} + \mathcal{E}_{pl}^{(p)} \\
 \mathcal{H}^{(m)} &= \mathcal{E}_k^{(m)} + \mathcal{E}_{coll}^{(m)} \\
 \mathcal{H}^{(b)} &= \mathcal{E}_{pl}^{(b)} \\
 \mathcal{E}_k^{(c)} &= \frac{\rho_c A}{2} \|\partial_t w\|_{\mathcal{D}}^2 + \frac{\rho_c A}{2} \|\partial_t \zeta\|_{\mathcal{D}}^2 \\
 \mathcal{E}_{pl}^{(c)} &= \frac{T_0}{2} \|\partial_x w\|_{\mathcal{D}}^2 + \frac{EI}{2} \|\partial_x^2 w\|_{\mathcal{D}}^2 + \frac{T_0}{2} \|\partial_x \zeta\|_{\mathcal{D}}^2 \\
 \mathcal{E}_{pnl}^{(c)} &= \frac{EA - T_0}{2} \left\| \frac{(\partial_x w)^2}{2} + \partial_x \zeta \right\|_{\mathcal{D}}^2 \\
 \mathcal{E}_k^{(p)} &= \frac{\rho_p H}{2} \|\partial_t W\|_{\mathcal{A}}^2 \\
 \mathcal{E}_{pl}^{(p)} &= \frac{D_x}{2} \|\partial_x^2 W\|_{\mathcal{A}}^2 + \frac{D_y}{2} \|\partial_y^2 W\|_{\mathcal{A}}^2 + D_{xy} \|\partial_y \partial_x W\|_{\mathcal{A}}^2 \\
 \mathcal{E}_k^{(m)} &= \frac{M_m}{2} (\partial_t U_m)^2 \\
 \mathcal{E}_{coll}^{(m)} &= \phi(\eta_m) \\
 \mathcal{E}_{pl}^{(b)} &= \frac{K_t}{2} (w(L) - W(\mathbf{x}_b))^2 + \frac{K_l}{2} (\zeta(L) - W(\mathbf{x}_b))^2
 \end{aligned}$$

Per la validità di (20), le condizioni al contorno per la corda e per la tavola devono essere passive. Anche la connessione tra la tavola armonica e il capo della corda deve essere effettuata in modo conservativo. In questo lavoro, le seguenti condizioni al contorno saranno utilizzate

$$w(0) = \zeta(0) = \partial_x^2 w(0) = \partial_x^2(L) = 0 \quad (21a)$$

$$W = \partial_n^2 W = 0 \quad \text{lungo il bordo della tavola} \quad (21b)$$

dove  $n$  indica la direzione normale al bordo della tavola. Queste condizioni al contorno sono completate da (14) (ovvero, la condizione al contorno per la corda, a  $x = L$ ), che garantiscono la passività del sistema. L'energia totale è non negativa (se  $EA \geq T_0$ ). La conservazione di un'energia non negativa, come noto, pone dei limiti superiori alla crescita delle norme delle soluzioni, e delle loro derivate [11].

### 3. IL MODELLO DISCRETO

In questa sezione, lo schema numerico utilizzato per la soluzione del sistema continuo verrà presentato. Lo schema si compone di una parte a differenze finite per la corda, i martelletti e gli strumenti di preparazione. La tavola armonica verrà risolta attraverso uno schema modale. Data la natura estremamente non lineare del sistema, sarà fondamentale il rispetto della passività del sistema discreto, ovvero di una legge di conservazione dell'energia del sistema discreto. Le condizioni per la non negatività dell'energia potranno essere interpretate come le condizioni di stabilità per lo schema numerico.

#### 3.1 Schema per le corde

Prima di procedere, è opportuno introdurre la notazione degli operatori a differenza finita usati per la corda, i mar-

telletti e gli strumenti di preparazione. La corda è definita su una griglia di punti. Le funzioni continue  $w(t, x)$ ,  $\zeta(t, x)$  sono rimpiazzate da funzioni definite sulla griglia, e denominate  $w_s^n$ ,  $\zeta_s^n$ , dove  $t = nk$ ,  $x = sh$ ;  $(n, s) \in \mathbb{N}$ ,  $0 \leq s \leq S$ ;  $k$  è il passo temporale (cioè l'inverso moltiplicativo della frequenza di campionamento), e  $h$  è la spaziatura della griglia. L'operatore di identità è definito come

$$1w_s^n = w_s^n$$

Gli operatori temporali sono definiti come

$$\begin{aligned}
 e_{t+} w^n &= w^{n+1}, & e_{t-} w^n &= w^{n-1} \\
 \delta_{t+} &\triangleq \frac{e_{t+} - 1}{k}, & \delta_{t-} &\triangleq \frac{1 - e_{t-}}{k} \\
 \delta_t &\triangleq \frac{e_{t+} - e_{t-}}{2k}, & \delta_{tt} &\triangleq \delta_{t+} \delta_{t-}
 \end{aligned}$$

Operatori di media temporale sono definiti come

$$\begin{aligned}
 \mu_{t+} \mu_{t-} &\triangleq \frac{1 + e_{t+}}{2}, & \mu_{t-} &\triangleq \frac{1 + e_{t-}}{2} \\
 \mu_t &\triangleq \frac{e_{t+} + e_{t-}}{2}, & \mu_{tt} &\triangleq \mu_{t+} \mu_{t-}
 \end{aligned}$$

Gli operatori spaziali sono definiti come

$$\begin{aligned}
 e_{x+} w_s &= w_{s+1}, & e_{x-} w_s &= w_{s-1} \\
 \delta_{x+} &\triangleq \frac{e_{x+} - 1}{h}, & \delta_{x-} &\triangleq \frac{1 - e_{x-}}{h}, & \delta_x &\triangleq \frac{e_{x+} - e_{x-}}{2h}
 \end{aligned}$$

Operatori temporali di secondo e quarto ordine sono definiti come

$$\delta_{xx} \triangleq \delta_{x+} \delta_{x-}, \quad \delta_{xxxx} \triangleq \delta_{xx} \delta_{xx}$$

Prodotto scalare e norma sono introdotti come

$$\langle a, b \rangle_{\mathbb{S}} \triangleq \sum_{s=0}^S h a_s b_s \quad (22a)$$

$$\|a\|_{\mathbb{S}}^2 \triangleq \sum_{s=0}^S h a_s^2 \quad (22b)$$

Infine, le distribuzioni temperate  $g_i$  in (23a) verranno sostituite da delta di Kroneker normalizzate, cioè

$$g_i = \frac{\delta_{i,s}}{h}$$

dove  $\delta_{i,s} = 1$  se e solo se  $i = s$ .

Con queste definizioni, il sistema (23) diventa

$$\mathfrak{L}(w) = \mathfrak{N}\mathfrak{L}(p, v) + \sum_i g_i \mathfrak{F}_i \quad (23a)$$

$$\mathfrak{G}(\zeta) = \mathfrak{N}\mathfrak{G}(v) \quad (23b)$$

$$v = \delta_{x-} w \quad (23c)$$

$$p = \delta_{x-} \zeta \quad (23d)$$

Gli operatori lineari sono definiti nella seguente maniera

$$\mathfrak{L}(w) = \rho_c A \delta_{tt} w - T_0 \delta_{xx} w + EI \delta_{xxxx} w \quad (24a)$$

$$\mathfrak{G}(\zeta) = \rho_c A \delta_{tt} \zeta - EA \delta_{xx} \mu_{tt} \zeta \quad (24b)$$

Gli operatori non lineari sono

$$\mathfrak{NL}(p, v) = \frac{EA - T_0}{2} \delta_{x+} [v^2 \mu_t v + 2v \mu_{tt} p] \quad (25a)$$

$$\mathfrak{NG}(v) = \frac{EA - T_0}{2} \delta_{x+} [v \mu_t v] \quad (25b)$$

Le forze al contorno  $s = S$  diventano

$$\mathfrak{F}_t = \mathfrak{F}_t^{(l)} + \mathfrak{F}_t^{(nl)} \quad (26a)$$

$$\mathfrak{F}_l = \mathfrak{F}_l^{(l)} + \mathfrak{F}_l^{(nl)} \quad (26b)$$

$$\mathfrak{F}_t^{(l)} = T_0 \delta_{x+} w - EI \delta_{x+} \delta_{xx} w$$

$$\mathfrak{F}_t^{(nl)} = \frac{EA - T_0}{2} [(\delta_{x+} w)^2 \mu_t (\delta_{x+} w) + 2(\delta_{x+} w) \mu_{tt} (\delta_{x+} \zeta)]$$

$$\mathfrak{F}_l^{(l)} = EA \left[ \frac{k^2}{4} \delta_{tt} + 1 \right] \delta_{x+} \zeta$$

$$\mathfrak{F}_l^{(nl)} = \frac{EA - T_0}{2} [(\delta_{x+} w) \mu_t (\delta_{x+} w)]$$

### 3.2 Schema per il martelletto e per gli strumenti di preparazione

Le collisioni del martelletto verranno simulate con l'ausilio di uno schema implicito, che dà origine a un'equazione algebrica non lineare. Questa dovrà essere risolta tramite un processo iterativo, qui Newton-Raphson. Lo schema è [9, 12]

$$\mathfrak{F}_m = \frac{\delta_{t+} (\mu_t - \phi(\eta_m))}{\delta_t \eta_m} \quad (27)$$

dove  $\phi(\eta_m)$  è definito in (9). Per gli strumenti di preparazione, espressioni simili sono utilizzate. Come detto, l'utilizzo di una tale forma per le forze di contatto permette di conservare un'energia discreta, e dunque di garantire la stabilità del sistema, anche in condizioni di forte non linearità. Per il martelletto, e per gli elementi mobili, l'equazione discreta del moto si scrive

$$M_m \delta_{tt} U_m = -\mathfrak{F}_m \quad (28)$$

### 3.3 Schema per la tavola armonica

Le equazioni per la tavola armonica saranno risolte con un approccio modale. Grazie a tale approccio, è possibile avere accesso ai tempi di decadimento in funzione della frequenza senza dover richiedere un ulteriore sforzo di calcolo. Inoltre, è possibile effettuare una riduzione del sistema, in maniera tale da mantenere un numero limitato di risonatori senza perdere eccessiva qualità sonora [13].

Pertanto, la soluzione  $W(\mathbf{x}, t)$  di (11a) è supposta tale che

$$W(\mathbf{x}, t) = \sum_{r=1}^R q_r(t) s_{r_x}(x) s_{r_y}(y) \quad (29a)$$

$$s_{r_x}(x) s_{r_y}(y) \triangleq \sin \frac{r_x \pi x}{L_x} \sin \frac{r_y \pi y}{L_y}, \quad (r_x, r_y) \in \mathbb{N} \quad (29b)$$

Utilizzando questa espressione, che soddisfa (21b), in (11), e proiettando sul modo  $r$  attraverso il prodotto scalare (13), si arriva all'equazione modale

$$\partial_t^2 q_r + \Omega_r^2 q_r = -\frac{4s_{r_x}(x_b) s_{r_y}(y_b)}{L_x L_y \rho_p H} (\mathcal{F}_t + \mathcal{F}_l) \quad (30)$$

dove si è assunto una forma tipo delta di Dirac per le distribuzioni temperate  $g_p^{(t)}, g_p^{(l)}$ . Le frequenze modali valgono

$$\Omega_r^2 = \frac{T_p}{\rho_p H} \left[ \frac{r_x^2 \pi^2}{L_x^2} + \frac{r_y^2 \pi^2}{L_y^2} \right] + \frac{D_x r_x^4 \pi^4}{\rho_p H L_x^4} + \frac{D_y r_y^4 \pi^4}{\rho_p H L_y^4} + \frac{2D_{xr} r_x^2 r_y^2 \pi^4}{\rho_p H L_x^2 L_y^2}$$

Uno schema risolutivo per (31) può essere dato in termini di operatori temporali a differenza finita, insieme all'espressione per le forze del ponticello (26). Perciò

$$\delta_{tt} q_r^n + \Omega_r^2 q_r^n = -\frac{4s_{r_x}(x_b) s_{r_y}(y_b)}{L_x L_y \rho_p H} (\mathfrak{F}_t + \mathfrak{F}_l) \quad (31)$$

### 3.4 Schema per il ponticello

Per il ponticello, lo schema numerico è

$$\mathfrak{F}_t = -K_t \left( w_S^n - \sum_r q_r^n s_{r_x}(x_b) s_{r_y}(y_b) \right) \quad (32a)$$

$$\mathfrak{F}_l = -K_l \left( \zeta_S^n - \sum_r q_r^n s_{r_x}(x_b) s_{r_y}(y_b) \right) \quad (32b)$$

dove sono stati esplicitati gli indici spazio-temporali delle funzioni di griglia, e delle coordinate modali. Si noti che le forze così definite sono esplicite nelle coordinate modali  $q_r$  (cioè, dipendono dal passo temporale  $n$ , e non da  $n+1$ ).

### 3.5 Conservazione dell'energia del sistema discreto

La controparte discreta di (19) si scrive dunque

$$\mathcal{L}(w) = \mathfrak{NL}(p, v) + g_m \mathfrak{F}_m \quad (33a)$$

$$\mathcal{G}(\zeta) = \mathfrak{NG}(v) \quad (33b)$$

$$v = \delta_{x-} w \quad (33c)$$

$$p = \delta_{x-} \zeta \quad (33d)$$

$$\delta_{tt} q_r^n + \Omega_r^2 q_r^n = -\frac{4s_{r_x}(x_b) s_{r_y}(y_b)}{L_x L_y \rho_p H} (\mathfrak{F}_t + \mathfrak{F}_l) \quad (33e)$$

$$M_m \delta_{tt} M_m = -\mathfrak{F}_m \quad (33f)$$

dove l'indice  $r$  in (33e) è intero e tale che  $r \in [1, R]$ . Per dimostrare l'esistenza di un hamiltoniano discreto, è sufficiente prendere un prodotto scalare tipo (22) di (33a) con  $\delta_t w$ , di (33b) con  $\delta_t \zeta$ . Si moltiplichino inoltre (33e) per  $\delta_t q_r$ , e (33f) per  $\delta_t U_m$ . Si sommino infine tutte le equazioni. Dopo somme per parti (non banali), e dopo l'utilizzo di non difficili, ma laboriose identità, è possibile arrivare al seguente bilancio energetico [14, 11]

$$\delta_{t+} \delta^n = 0 \quad (34)$$

dove

$$\begin{aligned}
 \mathfrak{H} &= \mathfrak{H}^{(c)} + \mathfrak{H}^{(p)} + \mathfrak{H}^{(m)} + \mathfrak{H}^{(b)} \\
 \mathfrak{H}^{(c)} &= \mathfrak{E}_k^{(c)} + \mathfrak{E}_{pl}^{(c)} + \mathfrak{E}_{pnl}^{(c)} \\
 \mathfrak{H}^{(p)} &= \mathfrak{E}_k^{(p)} + \mathfrak{E}_{pl}^{(p)} \\
 \mathfrak{H}^{(m)} &= \mathfrak{E}_k^{(m)} + \mathfrak{E}_{coll}^{(m)} \\
 \mathfrak{H}^{(b)} &= \mathfrak{E}_{pl,t}^{(b)} + \mathfrak{E}_{pl,l}^{(b)} \\
 \mathfrak{E}_k^{(c)} &= \frac{\rho_c A}{2} \|\delta_{t-} w^n\|_S^2 + \frac{\rho_c A}{2} \|\delta_{t-} \zeta^n\|_S^2 \\
 \mathfrak{E}_{pl}^{(c)} &= \frac{T_0}{2} \langle v^n, v^{n-1} \rangle_S^2 + \frac{EI}{2} \langle \delta_{xx} w^n, \delta_{xx} w^{n-1} \rangle_S^2 \\
 &\quad + \frac{T_0}{2} \|\mu_{t-} p^n\|_S^2 \\
 \mathfrak{E}_{pnl}^{(c)} &= \frac{EA - T_0}{2} \left\| \frac{(q^n)(q^{n-1})}{2} + \mu_{t-} p^n \right\|_S^2 \\
 \mathfrak{E}_k^{(p)} &= \frac{\rho_p H L_x L_y}{8} \sum_{r=1}^R (\delta_{t-} q_r^n)^2 \\
 \mathfrak{E}_{pl}^{(p)} &= \frac{\rho_p H L_x L_y}{8} \sum_{r=1}^R \Omega_r^2 (q_r^n)(q_r^{n-1}) \\
 \mathfrak{E}_k^{(m)} &= \frac{M_m}{2} (\delta_{t-} U_m^n)^2 \\
 \mathfrak{E}_{coll}^{(m)} &= \mu_{t-} \phi(\eta_m^n) \\
 \mathfrak{E}_{pl,t}^{(b)} &= \frac{K_t}{2} \left( w_S^n - \sum_r q_r^n s_{r_x}(x_b) s_{r_y}(y_b) \right) \times \\
 &\quad \left( w_S^{n-1} - \sum_r q_r^{n-1} s_{r_x}(x_b) s_{r_y}(y_b) \right) \\
 \mathfrak{E}_{pl,l}^{(b)} &= \frac{K_l}{2} \left( \zeta_S^n - \sum_r q_r^n s_{r_x}(x_b) s_{r_y}(y_b) \right) \times \\
 &\quad \left( \zeta_S^{n-1} - \sum_r q_r^{n-1} s_{r_x}(x_b) s_{r_y}(y_b) \right)
 \end{aligned}$$

L'energia cosí definita è una forma quadratica in  $(w, \zeta, q)^n$ ,  $(w, \zeta, q)^{n-1}$ . Sebbene (34) sia una controparte discreta di (20), la non negatività di  $\mathfrak{H}^{n-1/2}$  deve essere dimostrata. La situazione è resa complessa dai termini di accoppiamento tra sistema modale della tavola, e il sistema a differenze finite della corda, nelle espressioni di  $\mathfrak{E}_{pl,t}^{(b)}$ ,  $\mathfrak{E}_{pl,l}^{(b)}$ . Tuttavia, si noti che

$$\begin{aligned}
 \mathcal{E}_{pl,t}^{(b)} &= \frac{K_t}{2} \left[ (\mu_{t-}) \left( w_S^n - \sum_r q_r^n s_{r_x}(x_b) s_{r_y}(y_b) \right) \right]^2 \\
 &\quad - \frac{K_t k^2}{8} \left[ (\delta_{t-}) \left( w_S^n - \sum_r q_r^n s_{r_x}(x_b) s_{r_y}(y_b) \right) \right]^2
 \end{aligned}$$

e che pertanto

$$\mathcal{E}_{pl,t}^{(b)} \geq -\frac{K_t k^2}{8} \left[ (\delta_{t-}) \left( w_S^n - \sum_r q_r^n s_{r_x}(x_b) s_{r_y}(y_b) \right) \right]^2$$

Da questo, è possibile trovare un limite inferiore, tale che

$$\mathcal{E}_{pl,t}^{(b)} \geq \begin{cases} -\frac{K_t k^2}{2} (\delta_{t-} w_S^n)^2 \\ -\frac{K_t k^2}{2} (\sum_r q_r^n s_{r_x}(x_b) s_{r_y}(y_b))^2 \end{cases}$$

Il primo caso è valido quando  $|\delta_{t-} w_S^n| \geq |\sum_r q_r^n s_{r_x}(x_b) s_{r_y}(y_b)|$ , il secondo caso quando  $|\delta_{t-} w_S^n| < |\sum_r q_r^n s_{r_x}(x_b) s_{r_y}(y_b)|$ .

Procedendo in modo analogo per  $\mathcal{E}_{pl,l}^{(b)}$ , è dunque possibile trovare condizioni per la non negatività dell'energia totale imponendo le condizioni separatamente su

1. la componente trasversale del moto della corda
2. la componente longitudinale del moto della corda
3. la tavola armonica

Senza entrare nei dettagli di calcolo per mancanza di spazio, tali condizioni verranno ora elencate.

1. Per la componente trasversale del moto, (cioè per le componenti dell'energia che includono solo la funzione di griglia  $w$ ), la condizione *necessaria e sufficiente* per la non negatività è

$$\rho_c A h^4 - K_t k^2 h^3 - T_0 k^2 h^2 - 4EI k^2 \geq 0 \quad (35)$$

Tale equazione va risolta numericamente, ad esempio tramite un processo iterativo tipo Newton-Raphson. È possibile dimostrare che una e una sola soluzione per  $h > 0$  esiste per tutti i valori di  $K_t > 0$ , ma tale dimostrazione verrà omessa.

2. Per la componente longitudinale del moto (cioè per le componenti dell'energia che includono solo la funzione di griglia  $w$ ), la condizione *necessaria e sufficiente* per la non negatività è

$$K_l < \frac{\rho_c A h}{k^2} \quad (36)$$

Si noti in particolare che, rispetto a schemi precedenti (ad esempio in [14]) per la corda non lineare, sia  $w$  che  $\zeta$  sono definite sulla stessa griglia, senza dover interpolare tra due griglie con spaziature diverse. Ciò ha un effetto benefico duplice: riduce il tempo di calcolo, e non pone problemi da sottocampionamento spaziale.

3. Per la tavola armonica, una condizione *sufficiente* per la non negatività dell'energia associata è

$$\frac{\Omega_R^2}{4} + \frac{4R \max(K_t, K_l)}{L_x L_y \rho_p H} \leq \frac{1}{k^2} \quad (37)$$

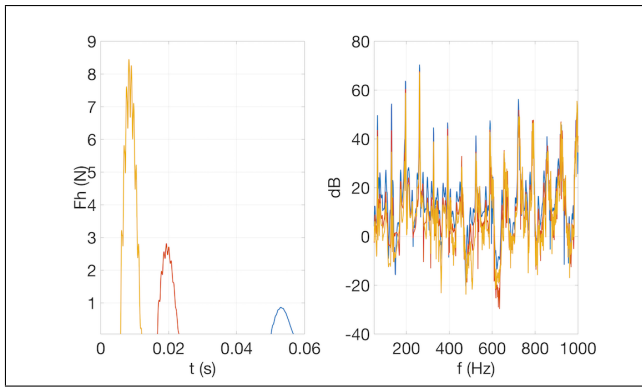
Si noti che, nel caso in cui  $K_t = K_l = 0$ , tale condizione si riduce alla condizione necessaria e sufficiente per l'oscillatore armonico [11].

Grazie a queste tre condizioni, è possibile trovare un limite inferiore per  $h$ , e un limite superiore per  $K_l$  e per il numero di modi  $R$ , una volta fissati  $k$  e  $K_t$ .

#### 4. ESEMPI NUMERICI

Ora che le condizioni di stabilità sono state enunciate, è possibile procedere alla descrizione di alcuni semplici esempi<sup>2</sup>. Gli esempi riportati hanno carattere esplicativo: lo

<sup>2</sup>Questi ed altri esempi possono essere ascoltati nella sezione *Prepared Piano* al seguente indirizzo: [mdphys.org](http://mdphys.org)

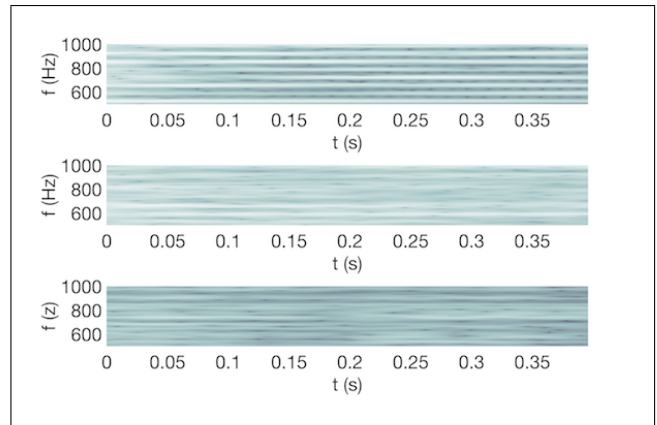


**Figure 1.** Simulazione di un D02 realistico. Lo strumento è composto da una sola corda di acciaio, di lunghezza  $L = 1.602$  m, raggio  $r_0 = 0.426$  mm, e di tensione uniforme  $T_0 = 200$  N. La tavola armonica è in legno, di lati  $L_x = 1.7$  m e  $L_y = 1.3$  m, e spessore  $H = 10$  mm. Una tensione uniforme di 200 N è applicata ai lati della tavola. Le rigidità del ponticello sono  $K_t = 5$  N/m,  $K_l = 10^4$  N/m. Il martelletto agisce a  $l_m = 0.05L$ , ha una massa di 10 grammi, un esponente  $p_m = 2.2$  e un coefficiente di rigidità  $K_m = 3.2 \cdot 10^8$ . In figura, vengono comparati i casi di velocità iniziale del martelletto pari a 0.2 m/s, 0.6 m/s, 1.7 m/s. La figura a sinistra rappresenta le forze di contatto nel corso del tempo per i tre casi, la figura di destra gli spettri. I segnali temporali sono stati registrati su un punto della tavola armonica.

strumento può (e deve) essere progettato in maniera complessa per essere sfruttato al meglio. Si noti inoltre che, sebbene la discussione fino ad ora sia stata svolta nel caso conservativo, gli esempi seguenti sono stati realizzati necessariamente con perdite, sia nelle corde, che negli strumenti di preparazione.

Il primo esempio è fornito in Figura 1. In tale esempio, uno strumento monocorda viene percosso a velocità crescenti. Si noti come, al crescere della velocità, il martelletto e la corda entrino in contatto con maggiore frequenza. Questo contribuisce ad arricchire lo spettro nelle alte frequenze.

Il secondo esempio è descritto nelle Figure 2 e 3. In queste figure, lo stesso strumento monocorda viene preparato con due preparazioni differenti: un sonaglio in un caso, e un sonaglio con ammortizzatore nel secondo caso, quest'ultimo lanciato ad una velocità iniziale estremamente elevata. Gli spettrogrammi della Figura 2 mostrano che, con una preparazione leggera, lo spettro viene leggermente modulato, lasciando le frequenze del D02 distinguibili nello spettro. In Figura 3 vengono proposti i segnali temporali dei tre casi, scalati in modo da poter essere rappresentati sulla stessa figura. A destra, gli spettri mostrano chiaramente la transizione da un caso dove i picchi di frequenza sono dettati dalla corda, a un caso dove essi rispondono alla dinamica degli strumenti di preparazione. Il moto caotico del sistema nel caso della preparazione con sonaglio e ammortizzatore rende la resa sonora imprevedibile ed efficace. Si tratta di un esempio di grande impatto, oltre che irrealizzabile in un vero pianoforte: la sua con-



**Figure 2.** Dall'alto verso il basso sono messi in evidenza gli spettrogrammi del pianoforte non preparato (in alto), del pianoforte con sonaglio (al centro), e del pianoforte con sonaglio e ammortizzatore (in basso). I parametri della tavola, della corda e del martelletto sono gli stessi della Figura 1. Il sonaglio per la figura al centro pesa 10 grammi, ha un esponente  $p_s = 2.5$  e un coefficiente di rigidità  $K_s = 10^7$ . Per la figura in basso, il sonaglio con ammortizzatore ha lo stesso peso e lo stesso esponente del precedente, ma un coefficiente di rigidità  $K_s = 10^3$ . Il coefficiente di perdita è  $\beta_a = 8 \cdot 10^{-4} K_s$  ed esponente  $p_a = p_s$ . L'ammortizzatore viene lanciato verso la corda alla velocità  $v_a = 100$  m/s.

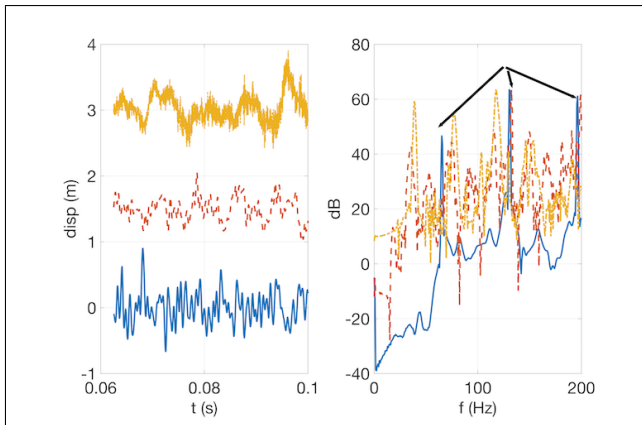
troparte digitale descritta in questo articolo risulta pertanto assai appetibile.

## 5. CONCLUSIONI

In questo lavoro, è stato presentato un modello per pianoforte preparato. Per le corde, è stato utilizzato il modello non lineare di Morse e Ingard, ed un nuovo schema a differenze finite è stato fornito per la soluzione delle equazioni. Differenze finite sono state usate per risolvere i contatti con i martelletti e gli strumenti di preparazione. Un approccio modale è stato invece utilizzato per la tavola armonica. Visto l'alto grado di non linearità del sistema completo, una dettagliata discussione sulla stabilità del sistema è stata fornita. La stabilità è stata fatta discendere dalla conservazione di un hamiltoniano discreto. In particolare, condizioni sufficienti per la stabilità sono state fornite indipendentemente per i moti trasversale e longitudinale della corda, e per il moto della tavola armonica. Infine, degli esempi numerici sono stati offerti per dimostrare alcune delle caratteristiche dello strumento digitale.

## 6. REFERENCES

- [1] R. Burger, *The Well-Prepared Piano*. San Pedro, CA, USA: Litoral Art Press, 1981.
- [2] J. Smith, "Physical modeling using digital waveguides," *Comput. Music J.*, vol. 16, no. 4, pp. 74–91, 1992.



**Figure 3.** Segnali temporali e spettri di frequenza per gli stessi tre casi della Figura 2. Le frecce nel diagramma a destra indicano le armoniche del DO2.

- [3] J. Chabassier, A. Chaigne, and P. Joly, “Modeling and simulation of a grand piano,” *J. Acoust. Soc. Am.*, vol. 134, no. 1, pp. 648–665, 2013.
- [4] M. Ducceschi and S. Bilbao, “Linear stiff string vibrations in musical acoustics: Assessment and comparison of models,” *J. Acoust. Soc. Am.*, vol. 140, no. 4, pp. 2445–2454, 2016.
- [5] B. Bank and L. Sujbert, “Generation of longitudinal vibrations in piano strings: From physics to sound synthesis,” *J. Acoust. Soc. Am.*, vol. 117, no. 4, pp. 2268–2278, 2005.
- [6] P. Morse and U. Ingard, *Theoretical Acoustics*. Princeton, NJ, USA: Princeton University Press, 1968.
- [7] B. Bank, S. Zambon, and F. Fontana, “A modal-based real-time piano synthesizer,” *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 18, no. 4, pp. 809–821, 2010.
- [8] A. Stulov, “Experimental and computational studies of piano hammers,” *Acta Acust. Acust.*, vol. 91, pp. 1086—1097, 2005.
- [9] S. Bilbao, A. Torin, and V. Chatziioannou, “Numerical modeling of collisions in musical instrument,” *Acta Acust. Acust.*, vol. 101, no. 1, pp. 155–173, 2015.
- [10] A. Chaigne, B. Cotte, and R. Viggiano, “Dynamical properties of piano soundboards,” *J. Acoust. Soc. Am.*, vol. 133, no. 4, pp. 2456–2466, 2013.
- [11] S. Bilbao, *Numerical Sound Synthesis: Finite Difference Schemes and Simulation in Musical Acoustics*. Chichester, UK: Wiley, 2009.
- [12] M. Ducceschi and S. Bilbao, “Modelling collisions of nonlinear strings against rigid barriers: Conservative finite difference schemes with application to sound synthesis,” in *Proc. ICA 2016*, (Buenos Aires, Argentina), September 2016.
- [13] M. Ducceschi and C. Webb, “Plate reverberation: Towards the development of a physical model for the working musician,” in *Proc. ICA 2016*, (Buenos Aires, Argentina), September 2016.
- [14] S. Bilbao, “Conservative numerical methods for nonlinear strings,” *J. Acoust. Soc. Am.*, vol. 118, p. 3316, 2005.

# MISURE DI HRTF SU UNA TESTA KEMAR CON PADIGLIONI AURICOLARI INTERCAMBIABILI

**Simone Spagnol**

Dip. di Ingegneria dell'Informazione  
 Università degli Studi di Padova  
 Padova, Italia  
 spagnols@dei.unipd.it

**Kristján Bjarki Purkhús, Sverrir Karl Björnsson, Rúnar Unnthórsson**

School of Engineering and Natural Sciences  
 University of Iceland  
 Reykjavík, Islanda  
 runson@hi.is

## SOMMARIO

Questo articolo ha come oggetto l'acquisizione di un insieme di misure di Head-Related Transfer Function (HRTF) sviluppata presso la University of Iceland. Le HRTF, campionate su una griglia sferica di 1513 posizioni, sono state misurate con una testa KEMAR dotata di volta in volta di uno tra 20 diversi padiglioni auricolari sinistri artificiali, ottenuti artigianalmente da accurati calchi di teste umane. Le valutazioni preliminari qui contenute suggeriscono la robustezza e la riproducibilità delle misure acustiche, così come la loro fedeltà rispetto alle HRTF KEMAR di riferimento. L'obiettivo della presente raccolta dati è di fornire un input affidabile per metodologie di determinazione di pattern spettrali individuali a partire da rappresentazioni 3D dell'orecchio e, più in generale, di incoraggiare studi sulla relazione tra HRTF e dati antropometrici tramite tecniche di apprendimento automatico.

## 1. INTRODUZIONE

La maggior parte delle tecniche per la resa binaurale del suono fa affidamento sull'utilizzo di Head-Related Transfer Function (HRTF), ovvero filtri che catturano gli effetti acustici del corpo umano [1]. L'uso di HRTF permette una simulazione realistica del segnale che giunge all'ingresso del canale uditivo in funzione della posizione della sorgente sonora nello spazio (azimut, elevazione e distanza). La resa ideale, ovvero quella che meglio approssima l'ascolto spaziale reale, richiede l'utilizzo di HRTF individuali misurate acusticamente sull'utente. Elaborando un segnale arbitrario tramite una coppia di HRTF individuali, una per canale, l'utente può fruire di autentiche esperienze sonore tridimensionali [2].

Sfortunatamente, ottenere HRTF individuali accurate comporta tutt'oggi avere accesso a strumentazione di laboratorio generalmente costosa e applicare procedure di registrazione invasive e/o faticose per l'utente [3]. Tale è la ragione per cui in pratica viene spesso favorito l'utilizzo di HRTF non individuali, misurate su manichini antropomorfi o su individui terzi. Diversi set di HRTF sono disponibili online; di questi, alcuni sono misurati su teste artificiali come

la KEMAR<sup>1</sup> mentre altri sono inclusi in database di misure individuali [4]. Molti database sono stati recentemente unificati in un formato comune di HRTF conosciuto come *Spatially Oriented Format for Acoustics* (SOFA).<sup>2</sup>

L'ovvio svantaggio delle HRTF non individuali è che difficilmente sono relazionabili all'antropometria dell'utente, in particolare quella del padiglione auricolare (detto anche *pinna*, termine che utilizzeremo in questo articolo). Nel momento in cui vengono usate per il rendering binaurale del suono, ciò si traduce in probabili e frequenti errori di localizzazione, quali inversioni (*front/back confusion*), errata percezione dell'elevazione, e localizzazione all'interno della testa [5]. Nonostante adattarsi a una coppia di HRTF non individuali sia possibile, questo processo richiede un tempo di apprendimento significativo [6].

Per tali ragioni, lo studio della relazione tra caratteristiche della pinna (forma, dimensioni, orientazione) ed HRTF rappresenta un passo fondamentale verso la comprensione dei meccanismi acustici sottintesi. Ciononostante, la precedente letteratura sulla relazione tra effetti acustici ed antropometria - tra cui applicazioni di metodi di regressione su database di HRTF [7, 8, 9, 10] - hanno prodotto risultati controversi, evidenziando come molte di queste relazioni non siano tuttora comprese appieno. Una delle possibili cause di tali ambiguità può essere individuata nel fattore umano, intendendo con questo problematiche quali il difficile controllo del posizionamento relativo tra microfono e orecchio reale, possibili movimenti del soggetto durante le misurazioni, e condizioni di misura non ideali. Tali problematiche possono dare luogo a misure poco accurate e non riproducibili, come nel caso di un precedente lavoro del primo autore riguardante la misurazione di HRTF su un campione di 25 soggetti umani [11]. D'altro canto, anche l'utilizzo di teste artificiali può dare adito ad errori: misure effettuate sulla stessa testa artificiale, infatti, possono addirittura risultare in set di HRTF molto diversi tra loro, a seconda del sistema di misurazione utilizzato [12].

In questo articolo presentiamo il design, l'implementazione e l'analisi preliminare di un nuovo set di misure di HRTF acquisite con una testa KEMAR dotata di pinne intercambiabili, con l'obiettivo di fornire un input affidabile per metodologie di determinazione di pattern spettrali individuali a partire da rappresentazioni 3D dell'utente [13]. In particolare, descriviamo la messa a punto un procedimento

Copyright: ©2018 Simone Spagnol et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution License 3.0 Unported](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

<sup>1</sup> <https://www.gras.dk/>

<sup>2</sup> <http://www.sofaconventions.org/>



**Figura 1.** Il sistema di misurazione di HRTF utilizzato.

artigianale per creare pinne artificiali per la testa KEMAR da calchi di teste umane, poi applicato a un campione di 20 pinne sinistre. Queste sono quindi state utilizzate come unica variabile antropometrica di una serie di misure di HRTF. Pur limitati dalle risorse disponibili, abbiamo progettato il setup di misurazione per essere quanto più accurato e stabile possibile, al fine di garantire affidabilità e riproducibilità delle relative misure acustiche.

## 2. METODOLOGIA

### 2.1 Sistema di misurazione

Le HRTF presentate in questo lavoro sono state acquisite automaticamente tramite il sistema di misurazione rappresentato in Figura 1. Esso consiste di un manichino KEMAR fissato a un supporto cilindrico rotante e di un altoparlante Genelec 8020CPM-6 fissato a un braccio rotante a forma di L. La configurazione KEMAR utilizzata è la 45BB-4 con pinne antropometriche standard (durezza 35 Shore-OO, GRAS KB5000/KB5001) e microfoni a pressione da mezzo pollice (GRAS 40AO) situati all'ingresso dei canali. I due canali audio sono collegati a una scheda RME Fireface 802 collegata a sua volta a un PC che esegue il software MATLAB.

La distanza tra il tweeter dell'altoparlante e il centro della testa è stata fissata alla costante di 1 m, indipendentemente dall'orientazione relativa di manichino e braccio rotante. In fase di misurazione, tale valore di distanza garantisce con sufficiente accuratezza la rappresentazione degli indicatori spettrali di localizzazione nel campo lontano [14]. Grazie all'utilizzo di un sistema di raggi laser indirizzati ai microfoni e alla punta del naso della KEMAR, è stato possibile ottenere un corretto allineamento del sistema di misurazione sui tre assi. Tali scelte sperimentali permettono di ottenere una collezione di HRTF in un range che copre tutti gli azimut ed elevazioni che vanno da  $-50^\circ$  a  $90^\circ$ , secondo un sistema di coordinate polari verticali.

La rotazione del manichino e del braccio è gestita indipendentemente da due motori passo-passo (JVL Industri Elektronik MST001A) ad alta coppia (1.2 Nm), controllati attraverso due azionatori digitali (Geckodrive G213V) in modalità full-step (200 passi per rivoluzione, singolo passo di  $1.8^\circ$ ). Al fine di aumentare la coppia e la risoluzione angolare del sistema, due riduttori 100 : 1 sono stati installati tra il primo motore e il braccio rotante e tra il secondo motore e il manichino. Di conseguenza, il passo minimo di entrambi i sistemi rotanti è di  $0.018^\circ$ . Inoltre, per ridurre la coppia necessaria sul braccio, un contrappeso di 22 kg è stato applicato all'estremità più corta in modo tale da bilanciare il peso dell'altoparlante e dell'altro segmento del braccio. La comunicazione tra gli azionatori e il PC è anch'essa gestita tramite MATLAB.

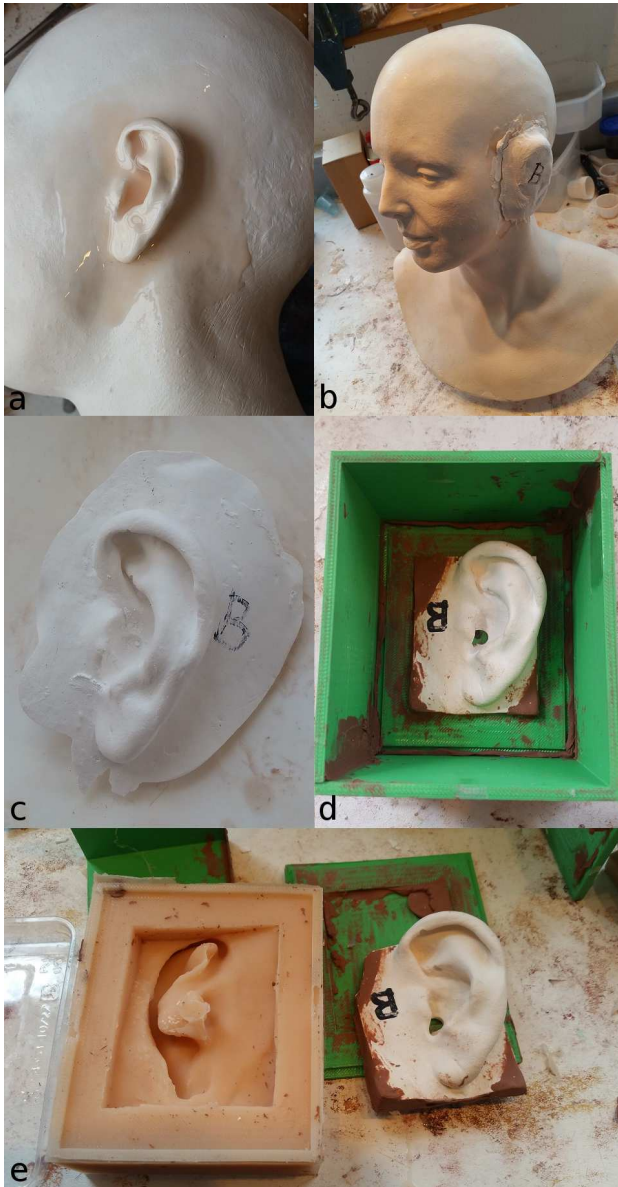
Per avere dei riferimenti assoluti di azimut ed elevazione durante le misurazioni, sono stati apposti rispettivamente un marcatore fisso nella parte inferiore del manichino (azimut  $0^\circ$ ) e una livella a bolla sul segmento più lungo del braccio (elevazione  $0^\circ$ ). Come accorgimento per evitare l'agrovigliarsi di cavi attorno al supporto del manichino e all'asse del braccio, ogni rivoluzione viene fatta seguire da una controrivoluzione in senso opposto.

### 2.2 Pinne artificiali

In questa sezione descriviamo la procedura artigianale utilizzata per creare le riproduzioni in silicone di pinne per la KEMAR. Al fine di avere una varietà sufficiente di forme per le misurazioni di HRTF, abbiamo applicato questa procedura alle pinne sinistre di 20 diversi soggetti. Tra questi, oltre alla KEMAR con pinne antropometriche standard, abbiamo considerato 19 riproduzioni in gesso di teste umane, concesse dal Saga Museum di Reykjavík. Tali teste, costruite a mano tra il 2001 e il 2003,<sup>3</sup> riproducono ad alta fedeltà i tratti antropometrici di 19 individui islandesi (7 femmine), di età compresa tra 7 e 77 anni al momento della costruzione. Si riportano di seguito le fasi principali del processo di riproduzione delle pinne.

*Step 1: Forma preliminare.* Applicare due strati di miscela di silicone all'orecchio della testa artificiale, il primo sottile con un pennello (Figura 2a) e il secondo più spesso (mezza tazza) con una spatola, lasciando asciugare ogni strato per un giorno prima dell'applicazione successiva. Onde ridurre la colatura su superfici verticali, aggiungere un paio di gocce di additivo tissotropico al silicone

<sup>3</sup> [https://www.youtube.com/watch?v=1102tc\\_bxKI](https://www.youtube.com/watch?v=1102tc_bxKI)



**Figura 2.** Alcune fasi del processo di costruzione di una forma di pinna per KEMAR.

dello strato spesso. Successivamente, applicare delle strisce di gesso bagnate sulla forma (Figura 2b) per agevolare la rimozione della forma stessa dalla testa artificiale e per evitare che si deformi nello step successivo.

*Step 2: Riproduzione in Jesmonite®.* Versare una miscela istantanea di mezza tazza d’acqua e due cucchiaini di Jesmonite® nella forma ottenuta nello Step 1, precedentemente posta sopra un tavolo vibrante (onde evitare la formazione di bolle d’aria), e lasciare indurire per 1–2 giorni. Successivamente, rimuovere la pinna in Jesmonite® (Figura 2c) e riempire eventuali fori o irregolarità con dell’argilla. Infine, per adattare la pinna alla guida della KEMAR, forare con un trapano la parte più interna della conca (in corrispondenza del canale uditivo) e smussare la base della pinna con una levigatrice a nastro.

*Step 3: Forma definitiva.* Riporre la riproduzione in Jesmonite® in una scatola di plastica aperta superiormente e fissarla alla base della scatola con dell’argilla (Figura



**Figura 3.** Le 20 riproduzioni di pinne sinistre ottenute con la procedura artigianale.

ra 2d). Successivamente, versare 2 tazze di miscela di silicone all’interno della scatola, precedentemente posta sul tavolo vibrante, e lasciare asciugare per 1 giorno. Infine, rimuovere la scatola e la pinna in Jesmonite® (Figura 2e).

*Step 4: Riproduzione in silicone.* Versare della cera per auto, in qualità di agente distaccante, nelle parti interne della forma ottenuta nello Step 3, e lasciare asciugare per mezza giornata. Successivamente, versare 1/3 di tazza di una miscela di silicone di durezza 25 Shore-A nella forma, e lasciare asciugare per 1 giorno. Infine, rimuovere la pinna in silicone e tagliare le parti eccedenti sulla base con un coltello. In Figura 3 riportiamo le 20 riproduzioni di pinne sinistre ottenute attraverso la procedura di cui sopra.

### 2.3 Procedura di misurazione

Le misurazioni sono state condotte durante diversi giorni, al di fuori dell’orario di lavoro (16.00 – 24.00) al fine di ridurre l’inquinamento acustico ambientale, all’interno di una stanza-ufficio dell’edificio Tækningarður della University of Iceland (vedi Figura 1). Poiché non abbiamo misurato la risposta a campo aperto del sistema, le misurazioni grezze includono la risposta in frequenza dell’altoparlante e dei microfoni così come alcuni effetti della stanza (ad esempio le riflessioni sui muri e sul pavimento, comunque minimizzate in fase di post-processing come spiegato successivamente).

Abbiamo utilizzato il metodo dello sweep logaritmico per registrare le singole risposte acustiche [15]. Il segnale di ingresso, il cui livello è stato tenuto costante durante l’intero programma di misurazione, riproduce tutte le frequenze tra 20 Hz e 20 kHz in 1 s, con una frequenza di campionamento  $f_s = 48$  kHz. Il livello SPL medio a 1 kHz per uno stimolo frontale come da misurazione trami-

**Tabella 1.** Posizioni della sorgente sonora.

| Elevazioni [deg] | [-45,45] | [50,70] | [75,85] | 90  |
|------------------|----------|---------|---------|-----|
| Passo [deg]      | 5        | 15      | 45      | 360 |
| No. di azimut    | 72       | 24      | 8       | 1   |

te un fonometro di Classe 1 posto al centro del sistema di riferimento è di 82 dB.

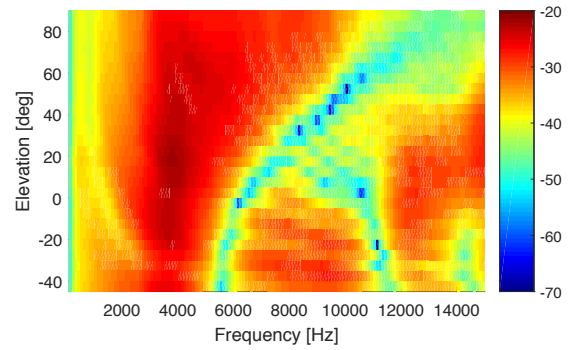
La posizione della sorgente sonora è univocamente determinata dall'azimut  $\theta$  e dall'elevazione  $\phi$  in coordinate polari verticali. Le elevazioni sono campionate uniformemente in passi da  $5^\circ$  tra  $-45^\circ$  e  $90^\circ$ , mentre gli azimut sono campionati diversamente a seconda dell'elevazione, come mostrato in Tabella 1. Questa scelta permette di evitare una concentrazione non necessaria di punti circostanti il polo superiore della sfera. Il numero totale di posizioni per misurazione è 1513.

Per ogni sessione di misurazione, le HRTF sono state registrate in ordine decrescente di elevazione, partendo da  $\phi = 90^\circ$ . Per ogni elevazione, il manichino è stato ruotato del corrispondente passo angolare fino a compiere un giro completo e le risposte allo sweep registrate ad ogni azimut. Alla fine del giro, il manichino è stato ruotato di  $360^\circ$  nel senso opposto e il braccio spostato all'elevazione successiva. Onde evitare la registrazione di rumori dovuti alla rotazione del sistema, delle pause di 0.5 s sono state inserite tra ogni operazione dei motori e la registrazione precedente/successiva. Date tali scelte sperimentali, la durata di una sessione di misurazione è di circa 105 minuti.

Ventitré sessioni di misurazione (3 misurazioni di controllo + 20 di test) sono state programmate in totale. Una pinna antropometrica destra standard (35 Shore-OO) è stata installata sul canale destro della KEMAR durante l'intero programma di misurazione, mentre la pinna sinistra cambiava ad ogni sessione. Nelle prime due sessioni di controllo (*CS1* and *CS2*) abbiamo installato rispettivamente la corrispondente pinna antropometrica sinistra e la sua variante di diversa durezza (55 Shore-OO). Queste sessioni sono state introdotte per controllare eventuali differenze tra materiali e per certificare la fedeltà della nostra riproduzione di pinna KEMAR alle originali. Le successive 20 sessioni sono state invece dedicate alle nostre 20 riproduzioni rappresentate in Figura 3, partendo dalla riproduzione KEMAR e procedendo con le 19 pinne umane (etichettate in ordine alfabetico da *A* a *S*). Nella terza e ultima sessione di controllo (*CS3*) abbiamo misurato di nuovo la KEMAR con entrambe le pinne standard, al fine di verificare la riproducibilità delle nostre misure. Tra ogni sessione di misurazione e la successiva, il sistema di misurazione è stato ricalibrato manualmente alla corretta posizione di partenza grazie ai riferimenti angolari assoluti.

## 2.4 Post-processing

L'estrazione delle HRTF dalle corrispondenti risposte allo sweep grezze è gestita da uno script di post-processing in MATLAB basato sui seguenti passi. Innanzitutto, identifichiamo il campione di partenza della risposta allo sweep tramite il calcolo della funzione di cross-correlazione  $\Psi[n]$



**Figura 4.** Risposte in ampiezza del set di HRTF *A*, piano mediano (pinna sx: riproduzione, pinna dx: KEMAR 35 Shore-OO).

tra la risposta allo sweep  $y[m]$  e il segnale d'ingresso  $s[m]$ . In particolare, troviamo il valore massimo di correlazione  $\Psi_M$  in  $\Psi[n]$  ed estraiamo il ritardo  $n_i$  del primo campione in  $\Psi[n]$  tale che  $\Psi[n] > 0.5\Psi_M$ . Quindi, il campione di partenza di  $y[m]$  corrisponde al ritardo  $n_j$  con valore massimo di correlazione all'interno di un intervallo di 21 campioni centrato in  $\Psi(n_i)$ . Questo escamotage è stato introdotto onde evitare di estrarre direttamente  $\Psi_M$  come campione di partenza che, a causa delle condizioni di misura ecoiche, potrebbe corrispondere a una riflessione sulla parete, specialmente nel caso di HRTF controlaterali.

Il passo successivo, coerentemente con il metodo dello sweep logaritmico, è un filtraggio inverso della risposta allo sweep al fine di ottenere la corrispondente risposta all'impulso. Calcoliamo dunque l'inverso  $S^{-1}[n]$  dello spettro di riferimento del segnale d'ingresso e applichiamo dei filtri di Butterworth passa-basso e passa-alto del secondo ordine per compensare l'originale assenza di contenuto sotto i 20 Hz e sopra i 20 kHz nel segnale d'ingresso. Quindi, la risposta all'impulso è ottenuta come

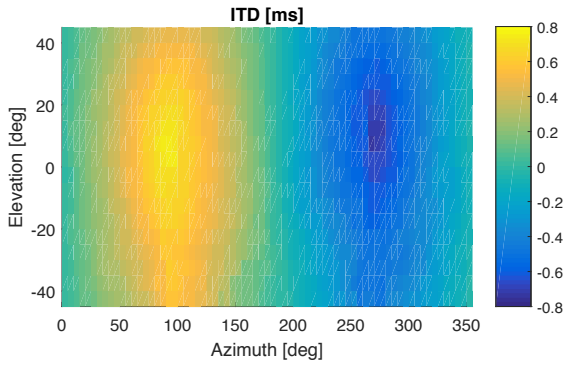
$$h[n] = \Re(\mathcal{F}^{-1}(\mathcal{F}(y[n]) * S^{-1}[n])), \quad (1)$$

dove  $\mathcal{F}$  e  $\mathcal{F}^{-1}$  sono la DFT e la DFT inversa.

Successivamente, ogni risposta all'impulso  $h[n]$  viene moltiplicata per una finestra di Hann da 128 campioni, allo scopo di rimuovere i contributi di tutte le onde riflesse sul sistema di misurazione e all'interno della stanza registrate approssimativamente oltre i 2.5 ms dal campione di partenza, ottenendo il segnale  $h_w[n]$ . La HRTF viene quindi infine semplicemente calcolata come risposta in ampiezza della DFT di  $h_w[n]$ . Come ultimo passo, la HRTF viene sottocampionata di un fattore 100 per ridurre la notevole dimensione dei dati (da 24000 a 240 campioni per HRTF).

## 3. RISULTATI

In questa sezione riportiamo alcuni risultati preliminari come validazione della fedeltà e riproducibilità del set di HRTF raccolto. Nella Figura 4 viene riportato un esempio di risposta in ampiezza delle HRTF sul piano mediano per una delle pinne artificiali (set *A*). Possiamo riconoscere degli effetti notevoli, quali l'avvallamento tra 1 e 2 kHz dovuto

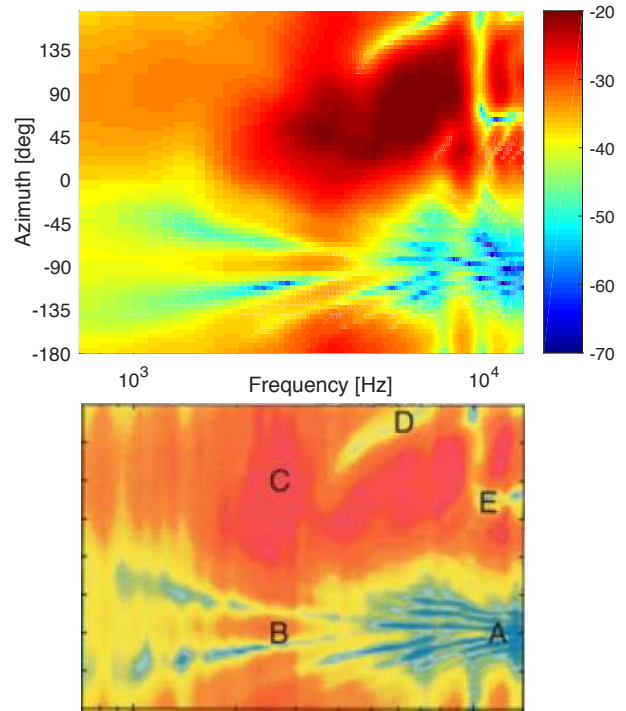


**Figura 5.** ITD del set di HRTF *A* (pinna sx: riproduzione, pinna dx: KEMAR 35 Shore-OO).

alla riflessione sulla spalla, la risonanza omnidirezionale tra 3 e 4 kHz e il pattern di picchi e notch dipendente dall'elevazione [1], con il più alto numero di notch registrato alle elevazioni inferiori [11]. La Figura 5 mostra le *interaural time differences* (ITD) per la stessa sessione di misurazione, calcolate come differenze temporali tra i campioni di partenza dei canali sinistro e destro di ogni risposta allo sweep. Coerentemente con la precedente letteratura [16], i valori di ITD variano tra zero (piano mediano) e  $\pm 0.8$  ms per le direzioni laterali. Si rileva inoltre l'asimmetria tra canale sinistro e destro, che risulta in massimi e minimi della ITD registrati a elevazioni differenti, dovuta alla presenza di due diverse pinne nella misurazione (pinna KEMAR e riproduzione da testa umana).

Per valutare la fedeltà delle nostre misure KEMAR, in Figura 6 riportiamo le misurazioni con la pinna di riferimento (35 Shore-OO, misurata nella sessione *CS1*) sul piano orizzontale e le paragoniamo con precedenti misurazioni della KEMAR alla distanza di 1 m ad opera di Brungart e Rabinowitz (Figura 5 in [17]). Se escludiamo la regione a più bassa frequenza, dove le differenze sono maggiori a causa della diversa configurazione del canale uditivo della KEMAR nei due setup sperimentali, possiamo riconoscere le stesse feature spettrali salienti sopra i 2 kHz, corrispondenti ai cinque punti di riferimento A-E (picchi e notch). Tale risultato qualitativo conferma l'efficacia del nostro setup di misurazione nel riprodurre ad alta fedeltà le misure KEMAR di riferimento.

Al fine di valutare la robustezza e la riproducibilità delle nostre misure, abbiamo calcolato la distorsione spettrale media nell'intervallo 3 – 10 kHz tra ogni set di HRTF e il set di riferimento KEMAR 35 Shore-OO. I risultati sono riportati in Tabella 2. Mentre per il canale destro si evidenzia soltanto rumore di misurazione, come dimostrato dalla bassa distorsione spettrale ( $< 0.5$  dB), i risultati del canale sinistro mostrano differenze notevoli tra gli effetti spettrali delle varie pinne. La distorsione spettrale media tra la pinna KEMAR e ognuna delle 19 riproduzioni di pinne umane è infatti sempre superiore ai 4.5 dB. Per contro, la riproduzione di pinna KEMAR fa registrare una distorsione media sensibilmente più bassa (inferiore a 1 dB), sorprendentemente inferiore anche a quella della pinna KEMAR originale di diversa durezza (55 Shore-OO)



**Figura 6.** Misurazioni KEMAR sul piano orizzontale nel presente lavoro (riquadro superiore) e in Brungart e Rabinowitz (riquadro inferiore, riprodotto da [17]).

**Tabella 2.** Distorsione spettrale media [dB] tra il set di HRTF di riferimento (*CS1*, KEMAR 35 Shore-OO) e tutti gli altri set (L: canale sx, R: canale dx).

| Set HRTF     | L      | R      |
|--------------|--------|--------|
| <i>CS2</i>   | 1.3736 | 0.1982 |
| <i>KEMAR</i> | 0.8407 | 0.2245 |
| <i>A</i>     | 4.5978 | 0.1885 |
| <i>B</i>     | 5.7394 | 0.2162 |
| <i>C</i>     | 6.2859 | 0.3389 |
| <i>D</i>     | 4.8688 | 0.1997 |
| <i>E</i>     | 5.7817 | 0.2291 |
| <i>F</i>     | 5.2341 | 0.2122 |
| <i>G</i>     | 5.1931 | 0.1656 |
| <i>H</i>     | 4.9951 | 0.3458 |
| <i>I</i>     | 5.6352 | 0.3494 |
| <i>J</i>     | 4.7819 | 0.1965 |
| <i>K</i>     | 5.6527 | 0.1990 |
| <i>L</i>     | 4.4466 | 0.2882 |
| <i>M</i>     | 4.7295 | 0.3364 |
| <i>N</i>     | 5.7147 | 0.2055 |
| <i>O</i>     | 5.0146 | 0.4611 |
| <i>P</i>     | 5.2498 | 0.2209 |
| <i>Q</i>     | 4.7395 | 0.2648 |
| <i>R</i>     | 6.5572 | 0.2806 |
| <i>S</i>     | 4.5860 | 0.1877 |
| <i>CS3</i>   | 0.7902 | 0.4186 |

misurata in *CS2*. Questo risultato suggerisce l'accuratezza della nostra procedura di riproduzione di pinne e l'impatto trascurabile della diversa durezza del silicone utilizzato. Infine, i risultati di *CS3* indicano che nonostante il riposizionamento della pinna KEMAR sul canale sinistro introduca della distorsione addizionale, questa non eccede il doppio del valore medio di distorsione dovuta al rumore di misurazione (canale destro).

#### 4. CONCLUSIONI

In questo articolo abbiamo presentato una collezione di HRTF di una testa KEMAR con pinne intercambiabili. Il setup e la procedura di misurazione sono stati descritti nel dettaglio, insieme alle operazioni di post-processing necessarie per estrarre le HRTF finali dalle risposte misurate. I nostri risultati preliminari certificano l'accuratezza e la varietà dei dati collezionati, che auspichiamo possano incoraggiare studi sulla relazione tra HRTF e dati antropometrici tramite tecniche di apprendimento automatico o altre recenti metodologie [18]. La collezione di HRTF, accompagnata da dettagliate scansioni e dati antropometrici delle 20 riproduzioni di pinne, verrà rilasciata online a breve e sarà scaricabile gratuitamente.

#### 5. RINGRAZIAMENTI

Gli autori vorrebbero ringraziare Ernst Backman, proprietario del Saga Museum di Reykjavík, per la sua preziosa assistenza nel processo di realizzazione delle pinne artificiali e per l'accesso al suo laboratorio e strumentazione.

#### 6. BIBLIOGRAFIA

- [1] C. I. Cheng and G. H. Wakefield, "Introduction to head-related transfer functions (HRTFs): Representations of HRTFs in time, frequency, and space," *J. Audio Eng. Soc.*, vol. 49, pp. 231–249, April 2001.
- [2] A. W. Bronkhorst, "Localization of real and virtual sound sources," *J. Acoust. Soc. Am.*, vol. 98, pp. 2542–2553, November 1995.
- [3] B. Xie, *Head-Related Transfer Function and Virtual Auditory Display*. Plantation, FL, USA: J.Ross Publishing, 2nd ed., June 2013.
- [4] V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano, "The CIPIC HRTF database," in *Proc. IEEE Work. Appl. Signal Process., Audio, Acoust.*, (New Paltz, New York, USA), pp. 1–4, October 2001.
- [5] H. Møller, M. F. Sørensen, C. B. Jensen, and D. Hammershøi, "Binaural technique: Do we need individual recordings?," *J. Audio Eng. Soc.*, vol. 44, pp. 451–469, June 1996.
- [6] C. Mendonça, "A review on auditory space adaptations to altered head-related cues," *Front. Neurosci.*, vol. 8, pp. 1–14, July 2014.
- [7] P. Bilinski, J. Ahrens, M. R. P. Thomas, I. J. Tashev, and J. C. Platt, "HRTF magnitude synthesis via sparse representation of anthropometric features," in *Proc. 39th IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP 2014)*, (Firenze, Italy), pp. 4501–4505, May 2014.
- [8] F. Grijalva, L. Martini, S. Goldenstein, and D. Florencio, "Anthropometric-based customization of head-related transfer functions using Isomap in the horizontal plane," in *Proc. 39th IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP 2014)*, (Firenze, Italy), pp. 4506–4510, May 2014.
- [9] S. Spagnol and F. Avanzini, "Frequency estimation of the first pinna notch in head-related transfer functions with a linear anthropometric model," in *Proc. 18th Int. Conf. Digital Audio Effects (DAFx-15)*, (Trondheim, Norway), pp. 231–236, December 2015.
- [10] S. Spagnol, S. Galesso, and F. Avanzini, "Stima di feature spettrali di HRTF mediante modelli antropometrici non lineari per la resa di audio 3D," in *Proc. XXI Colloquium on Music Informatics (XXI CIM)*, (Cagliari, Italy), pp. 129–135, October 2016.
- [11] S. Spagnol, M. Hiipakka, and V. Pulkki, "A single-azimuth pinna-related transfer function database," in *Proc. 14th Int. Conf. Digital Audio Effects (DAFx-11)*, (Paris, France), pp. 209–212, September 2011.
- [12] A. Andreopoulou, D. R. Begault, and B. F. G. Katz, "Inter-laboratory round robin HRTF measurement comparison," *IEEE J. Select. Topics Signal Process.*, vol. 9, pp. 895–906, August 2015.
- [13] S. Spagnol, M. Geronazzo, D. Rocchesso, and F. Avanzini, "Synthetic individual binaural audio delivery by pinna image processing," *Int. J. Pervasive Comput. Comm.*, vol. 10, pp. 239–254, July 2014.
- [14] S. Spagnol, E. Tavazzi, and F. Avanzini, "Distance rendering and perception of nearby virtual sound sources with a near-field filter model," *Appl. Acoust.*, vol. 115, pp. 61–73, January 2017.
- [15] S. Müller and P. Massarani, "Transfer-function measurement with sweeps," *J. Audio Eng. Soc.*, vol. 49, pp. 443–471, June 2001.
- [16] G. F. Kuhn, "Model for the interaural time differences in the azimuthal plane," *J. Acoust. Soc. Am.*, vol. 62, pp. 157–167, July 1977.
- [17] D. S. Brungart and W. M. Rabinowitz, "Auditory localization of nearby sources. Head-related transfer functions," *J. Acoust. Soc. Am.*, vol. 106, pp. 1465–1479, September 1999.
- [18] M. Geronazzo, S. Spagnol, and F. Avanzini, "A modular framework for the analysis and synthesis of head-related transfer functions," in *Proc. 134th Conv. Audio Eng. Soc.*, no. 8882, (Rome, Italy), May 2013.

# TECNOLOGIE PER L'INTERAZIONE SONORA IN CONTESTI DI REALTA' VIRTUALE E AUMENTATA IMMERSIVA

Michele Geronazzo

Department of Architecture, Design and Media Technologies  
Aalborg University Copenhagen  
mge@create.aau.dk

## ABSTRACT

Questo articolo espone le motivazioni, obiettivi e stati di avanzamento del progetto di ricerca *Acoustically-trained 3D audio models for virtual reality applications* sviluppato presso il Department of Architecture, Design & Media Technologies dell'Aalborg University Copenhagen, da me iniziato nel settembre 2017 della durata biennale. Lo sviluppo di tecnologie per il rendering multimodale accurato e realistico è centrale nel proporre strategie innovative in applicazioni che sfruttano l'interazione in tempo reale con e all'interno della realtà reale e virtuale (VR) o aumentata (AR). Questo progetto di ricerca contribuisce alla prossima generazione di framework per l'audio 3D che sarà in grado di estrarre informazioni acustiche proprie di utenti e ambiente circostante attraverso microfoni incorporati nelle cuffie consentendo: l'addestramento di (i) modelli parametrici per l'acustica dell'utente e della stanza, (ii) compensazione dinamica e trasparenza acustica della cuffia e (iii) definizione di nuove interazioni tra utente, dispositivo e ambiente. La fusione dell'audio virtuale/reale necessita di una piattaforma innovativa per una realtà acustica virtuale personalizzata. Verranno qui riassunte una serie di direzioni di ricerca attualmente in corso presso il *Multisensory Experience Lab*.

## 1. INTRODUZIONE

Il progetto *Acoustically-trained 3D audio models for virtual reality applications* trova le sue radici nelle discipline dell'informatica applicata all'interazione uomo-computer, alla modellazione dell'ascoltatore ed all'elaborazione del segnale acustico. In particolare, tutto ruota attorno alla spazializzazione del suono ha un elevato impatto in scenari applicativi come l'intrattenimento, ambienti condivisi virtuali/aumentati e nello studio dell'interazione umana nel mondo reale.<sup>1</sup>

L'obiettivo principale di questo programma di ricerca è la prototipazione e la valutazione sperimentale di un innovativo dispositivo audio in grado di rendere coerentemente

<sup>1</sup> Si vedano i recenti progetti europei: EPSRC EP/L000539/1, BEAMING – FP7-IP-248620 e TWO!EARS - FP7-STREP-618075

Copyright: ©2018 Michele Geronazzo et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution License 3.0 Unported](https://creativecommons.org/licenses/by/3.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

una realtà acustica virtuale personalizzata sovrainposta a scene acustiche reali. La interconnessione tra le informazioni acustiche individuali dell'ascoltatore raccolte nella *head-related transfer function* (HRTF) e le proprietà acustiche della stanza catturate da tale dispositivo è fondamentale ed è ancor oggi non adeguatamente trattata nella letteratura scientifica inibendo molte opportunità nel mercato globale del futuro.

Geronazzo e colleghi in [1, 2] hanno formalizzato una nuova metodologia di ricerca per la progettazione e la personalizzazione della HRTF sintetica, che prende il nome di approccio *mixed structural modeling* (MSM); il paradigma di modellazione strutturale [3] e le tecniche di selezione di HRTF generiche [4] informate da dati antropometrici individuali vengono combinate al fine di calcolare un modello MSM parametrizzato sulle caratteristiche dell'ascoltatore. Inoltre un mixer acustico di sorgenti sonore reali e virtuali deve tenere conto dei cambiamenti dinamici nell'ambiente [5]. Per questo un headset innovativo deve essere in grado di accedere a tutte queste informazioni per poter costruire un profilo utente contenente i parametri di rendering percettivamente rilevanti per il singolo ascoltatore [6], fornendo così un supporto alla personalizzazione per la riproduzione audio binaurale e le applicazioni realtà virtuale (VR) e aumentata (AR) [7].

Il framework di ricerca che è stato sviluppato consiste in tre tasselli fondamentali non ancora interconnessi tra di loro: (i) un motore di rendering audio completamente personalizzabile, (ii) un primo prototipo di dispositivo audio acusticamente trasparente e (iii) un ambiente di valutazione in ambiente VR. La piattaforma sviluppata fornisce una solida base per ulteriori valutazioni sperimentali e integrazioni con diversi dispositivi di realtà virtuale quali Samsung Gear VR, Oculus Rift e HTC Vive.

## 2. LAVORI CORRELATI

Le tecnologie per il rendering di audio spaziali (si veda [8] per una recente review sull'argomento) attraverso le cuffie devono essere in grado di modellare la *binaural room impulse response* (BRIR) che contiene le informazioni acustiche di una sorgente sonora in un specifico ambiente. La BRIR può essere suddivisa in due componenti: la risposta all'impulso della stanza (RIR), che definisce le proprietà acustiche di una stanza e la *head-related impulse response* (HRIR), che descrive acusticamente i contributi acustici dell'ascoltatore, comprendendo testa, orecchio esterno, tronco e spalle.

Per il rendering di una sorgente sonora virtuale, il segnale viene tipicamente filtrato dalla coppia di BRIR corrispondenti alla direzione spaziale della sorgente [9] e da una coppia di funzioni di trasferimento delle cuffie (HpTF) che tengono in considerazione le proprietà acustiche del dispositivo una volta indossato dall'ascoltatore [10]. Per fornire prestazioni di localizzazione del suono che siano equivalenti a quelle in campo libero sono necessarie HRTF e HpTF specifiche per l'ascoltatore [11]. Le HRTF sono solitamente misurate su un numero limitato di soggetti umani e teste artificiali facendo variare la posizione della sorgente sonora rispetto alla testa, mentre le HpTF guidano il processo di equalizzazione di diversi modelli di cuffia. La soluzione che meglio si avvicina alle condizioni di ascolto naturale prevede l'utilizzo di HRTF personalizzate e HpTF misurate sull'ascoltatore con una riverberazione artificiale percettivamente realistica [12, 13].

L'approccio MSM per la progettazione e la personalizzazione di HRTF sintetiche combina il paradigma di modellazione strutturale con tecniche di selezione di contributi non individualizzati [2]; una singola HRTF può venir sintetizzata come una combinazione di componenti strutturali, che possono venir scelte come componenti sintetiche o pre-registrate da misurazioni già disponibili. In entrambi i casi la personalizzazione si basa su dati antropometrici / acustici individuali, utilizzati per adattare i parametri del modello o per selezionare una componente all'interno di una serie di misurazioni disponibili. Per quanto riguarda la RIR, la letteratura scientifica propone numerosi modelli computazionale e riverberatori artificiali in grado di descrivere l'acustica di una stanza e con lo scopo di auralizzare dinamicamente un certo ambiente. Alcuni esempi ibridi sono la modellazione *image-source* e il *ray/beam-tracing* per la modellazione delle prime riflessioni assieme ad una gestione statistica della coda riverberante [14]. La parametrizzazione di tali modelli può venir effettuata da informazioni acustiche estratte dall'ambiente che possono essere considerate come una stima approssimativa della RIR reale.

A tale scopo, un approccio a filtraggio adattivo può essere in grado di estrarre HRIR / RIR al volo servendosi di movimenti casuali della testa [15]. L'estrazione di parametri acustici per la modellazione di HRIR / RIR può venir effettuata grazie al supporto *ear-through* in cuffie quali Roland CS-10EM e Philips SHN2500, che permette l'elaborazione delle informazioni ambientali estratte da microfoni o array microfonici incorporati negli auricolari. Questo tipo di input può consentire lo sviluppo di algoritmi per la localizzazione automatica della sorgente sonora, la separazione di sorgenti sonore e la fusione delle informazioni multimodali che circondano l'ascoltatore. Numerosi studi ed esempi applicativo possono venir rintracciati nel campo di ricerca della *robot audition* [16] e nello sviluppo di protesi acustiche [17]. Infine l'integrazione di motori di rendering per l'audio spaziale binaurale in tali dispositivi portatili permettono la fruizione personale di contenuti audio spazializzati durante lo spostamento o la guida, e senza rivolgersi verso una specifica fonte acustica magari al di fuori del campo visivo [18].

### 3. IL PROGETTO

Le attività di questo progetto possono essere suddivisa in tre direzioni di ricerca strettamente interconnesse.

#### 3.1 Tecniche e modelli per il rendering audio binaurale

Vengono prese in considerazione le principali componenti di un sistema per il rendering di audio binaurale in cuffia:

- numerosi criteri di selezione MSM grazie a studi approfonditi sulla connessione tra informazione acustica e antropometria [19, 20];
- algoritmi per l'equalizzazione e trasparentizzazione acustica di cuffie che siano in grado di considerare l'acustica del condotto uditivo, *leakage* e colorazione [21];
- posizionamento ottimale dei microfoni miniaturizzati per un rendering coerente della scena mista virtuale 7 reale per l'ascolto aumentato [22, 18]
- modellazione acustica della stanza e della sua dinamica attraverso l'analisi di BRIR misurate [14].

#### 3.2 Design del dispositivo audio

L'obiettivo principale di queste attività consiste nello sviluppo di algoritmi che siano in grado di stimare i parametri individuali dei modelli MSM, dell'accoppiamento acustico cuffia-ascoltatore in tempo reale e i parametri acustici della stanza attraverso tecniche di apprendimento automatico su segnali da array di microfoni incorporati nella cuffia come input. La direzione di arrivo (*direction-of-arrival* - DOA) e il rapporto suono diretto / riverberato sono in grado di fornire informazioni utili per alcuni di questi parametri. Un esempio di applicazione che segue questa idea è il concetto di *selfie acustico* sviluppata in [23]. Inoltre un'analisi computazionale della rappresentazione dell'informazione uditiva per l'ambiente circostante può permettere la separazione delle sorgenti sonore dalle proprietà acustiche della stanza [24].

Il prototipo in fase di sviluppo potrà essere integrato nei servizi Web per l'hosting, la calibrazione, l'adattamento e il recupero dei singoli parametri di rendering grazie all'esistente framework *HRTF On Demand for Binaural Audio* (HOBA) [25].

#### 3.3 Interazione e valutazione

La valutazione della personalizzazione, della trasparenza acustica, della coerenza negli indicatori spaziali e in generale della qualità dell'esperienza legata all'uso del dispositivo viene effettuata con metriche di errore legate al segnale, l'utilizzo di modelli uditivi [26, 27], e test di ascolto soggettivo con sorgenti sonore reali e virtuali in ambienti. In particolare, grazie all'inventario della qualità audio spaziale (SAQI) [28], e al vocabolario dei giudizi sulla qualità delle HRTF [29] si può effettuare una valutazione del grado di esternalizzazione e naturalezza dell'esperienza di ascolto. La definizione di specifici casi di studio servirà da banco di lavoro per testare il prototipo di dispositivo e una sua virtualizzazione. Le interazioni di base con l'ambiente reale/virtuale (ad esempio raggiungimento degli obiettivi,

riconoscimento degli oggetti/parlatori, ecc.) possono venir esplorate con una crescente complessità nelle condizioni testate fornendo un feedback per il miglioramento nella progettazione del dispositivo.

#### 4. LAVORI IN CORSO

In questa sezione vengono riassunti i principali risultati del progetto di ricerca dopo questo suo primo anno di svolgimento. Si possono evincere tre direzioni di ricerca principali che si sviluppano attorno ai seguenti settori: sviluppo di algoritmi per il rendering audio real-time, prototipazione e design di prodotto e interazione uomo-dispositivo. Sono state sviluppate separatamente in questo primo anno di progetto per poi prevederne la fusione durante il secondo ed ultimo anno di vita del progetto.

##### 4.1 Implementazione a modelli strutturali misti

Richiamando il formalismo dei modelli MSM, una *head-related transfer function parziale* ( $pHRTF$ ) contiene le informazioni acustiche misurate isolando specifiche parti del corpo (es. relative al contributo del solo orecchio esterno). Ogni componente strutturale può venir scelta all'interno di tre differenti raggruppamenti:

- componenti individuali (le  $pHRTF$  del soggetto  $i$ );
- componenti selezionate (le  $pHRTF$  di soggetti diversi da  $i$ );
- componenti modellate (le  $\widehat{pHRTF}$  sintetizzate).

L'approssimazione risultate ( $\widehat{HRTF}_I$ ) includerà  $S$  componenti selezionate,  $I$  componenti individuali e  $M$  componenti modellate:

$$\widehat{HRTF}_i = \bigotimes_{k=1}^S pHRTF_{s_k} \otimes \bigotimes_{k=1}^I pHRTF_{i_k} \otimes \bigotimes_{k=1}^M \widehat{pHRTF}_{m_k} \quad (1)$$

dove

$$i, s \in \mathcal{S}, \quad m \in \mathcal{M}$$

$$I + S + M = N$$

e  $\mathcal{S}$  e  $\mathcal{M}$  rappresentano le collezioni di soggetti e modelli per cui almeno una  $pHRTF$  o una  $\widehat{pHRTF}$  sia disponibile;  $\otimes$  è l'operatore che si riferisce ad una rappresentazione a filtri, e può identificare una connessione in serie o in parallelo;  $s_k$  e  $i_k$  indicano la  $k$ -esima componente parziale rispettivamente per un soggetto  $s$  e per il soggetto  $i$ ;  $m_k$  è la  $k$ -esima componente modellata.

Il MSM sviluppato per questo progetto prevede  $S = 1$ ,  $M = 2$ ,  $I = 0$ , e  $N = 3$ . La componente selezionata tiene in considerazione il contributo spettrale delle HRTF al di sopra dei 2 kHz e prevede la scelta della HRTF generica che possiede la distanza minima tra notch delle HRTF disponibili in uno specifico database e la stima dell'andamento dei notch dell'ascoltatore. Seguendo la metrica proposta nei seguenti studi [30, 20] l'andamento del primo notch nella HRTF viene predetto attraverso il tracciamento del contorno esterno dell'orecchio basandosi sulla foto laterale dell'orecchio dell'ascoltatore (v. esempio in Fig. 1). Mentre le componenti modellate vengono introdotte per tener in considerazione il contributo spettrale e temporale alle basse frequenze:

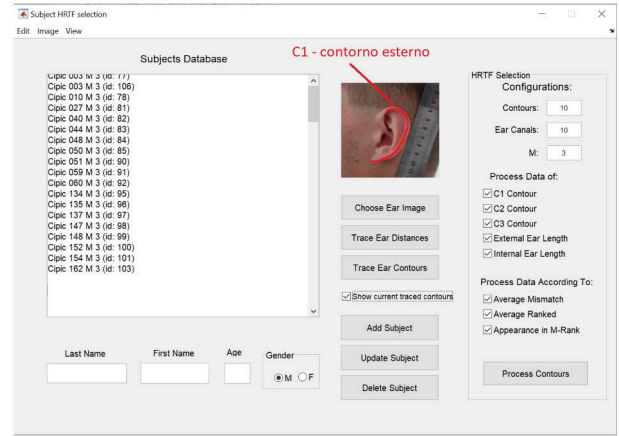


Figura 1: Software per il tracciamento contorni e selezione di HRTF.

- il contributo spettrale della testa al di sotto dei 2 kHz attraverso il filtro del primo ordine introdotto da Brown e Duda [3];
- l'*interaural time difference* (ITD) attraverso il modello sferico che considera un offset per la posizione delle orecchie rispetto all'asse inter-aurale della testa sferica [31];

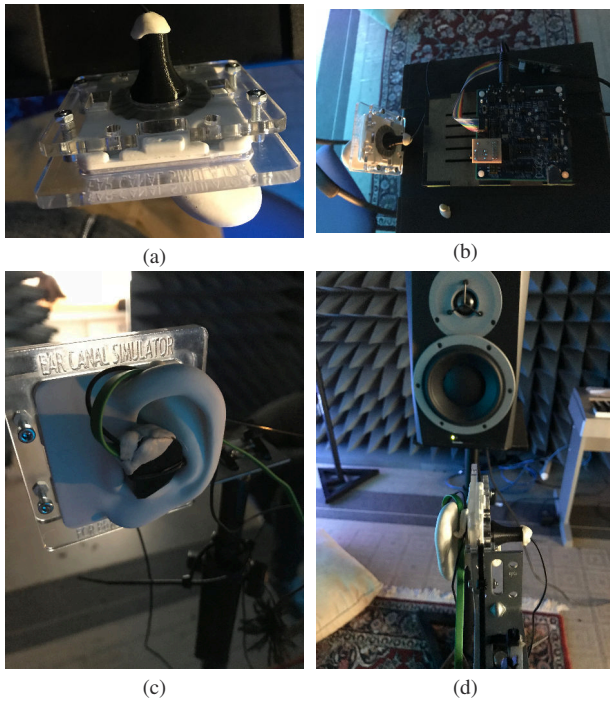
Per entrambe le componenti modellate la larghezza e lunghezza della testa dell'ascoltatore sono utilizzati come parametri nella formula di regressione lineare per il calcolo del raggio della testa sferica.

L'approccio fin qui descritto può venir considerato a tutti gli effetti una modellazione fisica dell'acustica dell'ascoltatore. Un ragionamento simile può venir effettuato per l'acustica dell'ambiente che circonda l'ascoltatore. Per questo progetto, si è deciso di adottare le *scattering dealy networks* [32] essendo un metodo computazionalmente efficiente e dall'alta flessibilità grazie all'esplicita definizione di parametri quali la dimensioni della stanza e i coefficienti di assorbimento delle pareti. L'algoritmo SDN esegue il rendering con precisione nel tempo e nello spazio considerando i percorsi del suono diretto (sorgente-ascoltatore) e le prime riflessioni calcolate secondo il metodo image-source per geometrie semplici o attraverso algoritmi di stima delle DOA e *powermap* basati su acquisizioni con array microfoniche [33]. Mentre le riflessioni di ordine superiore usano la stessa rete che congiunge le giunzioni delle prime riflessioni, l'ascoltatore e la sorgente, producendo così un'approssimazione sempre più grossolana dei punti di riflessione.

##### 4.2 Cuffie per la realtà acustica aumentata

Questo progetto ha tra i suoi obiettivi principali quello di fornire una piattaforma per il design di un nuovo dispositivo audio che sia in grado di potenziare l'uso delle cuffie che indossiamo ogni giorno, tipicamente in-ear, semplificandone la progettazione.

In accordo con Albrecht e colleghi [34] lo scenario applicativo di realtà aumentata audio necessita di alcune com-



**Figura 2:** Setup di test del dispositivo. (a) Orecchio artificiale/Simulatore con canale uditivo per le misurazioni, (b) ADAU1777 DSP e board prototipale, (c) Prototipo di cuffia montato sul simulatore, e (d) Posizionamento dell'altoparlante in relazione all'orecchio artificiale

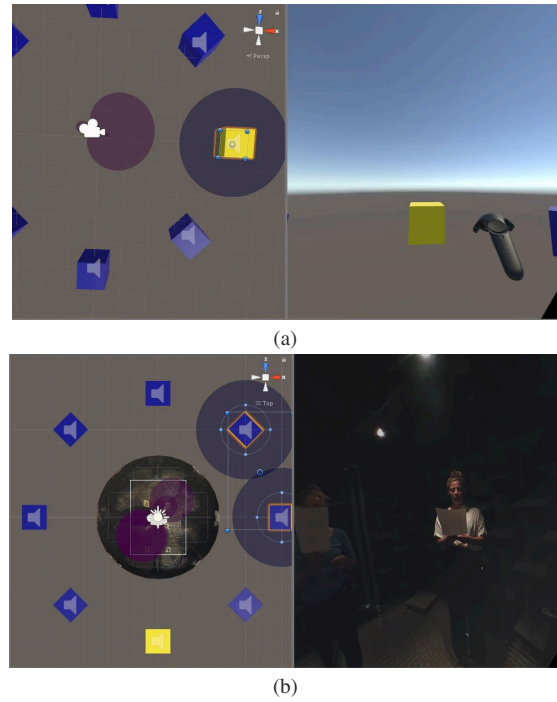
ponenti imprescindibili che possono venir riassunte ed estese nei seguenti punti:

- la cancellazione attiva del rumore (*active noise cancellation - ACN*) per isolare l'ascolto in cuffia;
- la compensazione dell'occlusione del canale uditivo per evitare l'amplificazione delle basse frequenze;
- la trasparenza acustica del dispositivo che consente di ascoltare in maniera naturale l'ambiente reale circostante;
- la de-riverberazione per l'estrazione delle caratteristiche acustiche delle sorgenti sonore;
- il *beamforming* per isolare acusticamente una direzione di ascolto preferita dall'utente;

La piattaforma è pensata e progettata seguendo i criteri di riproducibilità fornendo la documentazione necessaria e selezionando attrezzature disponibili alla portata della maggior parte dei ricercatori grazie a strutture quali il FabLab della Aalborg University o simili.<sup>2</sup> Il kit di sviluppo è altamente modulare *by design* per mantenere un elevato grado di personalizzazione e al momento contiene una collezione di algoritmi di base, un progetto di costruzione per l'alloggiamento di cuffia, driver ed elettronica di montaggio, e un progetto di elettronica PCB per il collegamento dei microfoni MEMS e del driver (v. Fig.2(b),(c) per delle foto del prototipo sviluppato).

Lo stato attuale della prototipazione delle cuffie in-ear consiste nello sviluppo di algoritmi per la compensazione dell'effetto occlusione e ANC utilizzando due disposi-

<sup>2</sup> <http://media.aau.dk/facilities/laboratories/electronics-fablab-workshop>



**Figura 3:** Scenari di valutazione per l'interazione dispositivo-ascoltatore in VR. (a) scenario astratto con feedback visivo minimale, e (b) scenario realistico con video 360. Entrambi gli scenari contengono 8 persone che parlano simultaneamente attorno all'ascoltatore.

tivi DSP modello ADAU1777 (v. Fig.2(b)) a bassa latenza ( $\approx 5$  ms). Le cuffie in-ear sono equipaggiate con microfoni MEMS interni ed esterni e sono state assemblate e testate acusticamente per un corretto funzionamento del dispositivo (v. Fig. 2(a),(d) per il setup di misurazione attraverso orecchio artificiale). Il collegamento elettronico tra DSP e microfoni MEMS è stato effettuato con un PCB aggiuntivo prodotta da una ditta di terze parti.

### 4.3 Interazione sonora e beamforming

L'esperienza di ascolto umano è caratterizzata dalla capacità di focalizzare l'attenzione in una determinata direzione, verso una specifica sorgente sonora [35]. L'attenzione uditiva selettiva si basa su un insieme di meccanismi cognitivi per determinare la DOA di un suono per creare una rappresentazione del percepito uditivo, rafforzato da un'esperienza intrinsecamente multimodale.

In questa progetto è importante comprendere come questi meccanismi influiscano nel comportamento dell'ascoltatore e fino a che punto possano essere supportati da dispositivi acustici che aggiungano un livello extra di artificialità come in apparecchi e protesi acustiche. L'obiettivo finale di questo insieme di attività è quello di definire quali interazioni in contesti di realtà aumentata acustica possano essere implementati come strumenti di supporto ad un *super uditivo*.

All'interno di questo progetto si sta sviluppando una piattaforma per la prototipazione virtuale di dispositivi acustici in cui è possibile testare diversi tipi di interazione in diverse simulazioni e situazioni. Al momento sono state

testate tre metafore di interazione in combinazione con algoritmi di *beamforming* per favorire l'attenzione selettiva del parlatore in contesti di AR:

- focus guidato dalla rotazione della testa;
- focus guidato dal movimento dell'occhio;
- utilizzo di un telecomando a distanza per il controllo dei parametri del dispositivo.

Sono state realizzate due scene in VR immersiva con HTC Vive in modo da testare il controllo dell'utente con e senza una virtualizzazione di dispositivo di ascolto aumentato. Un primo scenario rappresentato in Fig.3(a) con otto cubi virtuali disposti circolarmente attorno al partecipante e uno scenario realistico (Fig.3(b)) con otto persone reali in camera anecoica registrate attraverso un telecamera che acquisisce video a 360°. Gli utenti potevano controllare queste tre metafore agendo sui seguenti gradi di libertà: la direzione di focus, la larghezza del raggio di direzionalità e il livello di pressione sonora della sorgente sonora selezionata.

## 5. CONCLUSIONI

La realizzazione di un dispositivo che renda possibile una realtà aumentata acustica e un ascolto aumentato è un percorso ricco di sfide multidisciplinari. In particolare, la personalizzazione del rendering audio, il design del dispositivo acustico sono di vitale importanza per determinare il successo di tale ricerca. Questo progetto ha grandi potenzialità per quanto riguarda l'arricchimento delle interfacce utente attraverso feedback acustici più naturali e immersivi; si potrà dare così la possibilità a designer e ad aziende nel settore multimedia di sviluppare nuovi tipi di interazione e contenuti che mettano al centro l'informazione uditiva molto spesso messa in secondo piano rispetto a quella visiva.

Gli sviluppi futuri di tale progetto prevedono l'ottimizzazione del motore di rendering audio e la valutazione rigorosa della sua parametrizzazione. In particolare il testing di algoritmi per l'estrazione automatica dei parametri acustici della stanza e dell'ascoltatore tramite tecniche di *machine learning* and *machine audition* [36] sarà un tassello fondamentale. Infine il continuo sviluppo dell'ambiente VR di test per la fase di prototipazione del dispositivo, o meglio una sua virtualizzazione, sarà di vitale importanza per la scelta delle caratteristiche rilevanti nell'interazione tra ascoltatore e ambiente simulato aumentato.

## 6. RINGRAZIAMENTI

L'autore ringrazia i suoi studenti AAU Luis Vieira, Jason Tisseries, Baldur Kampmann, Antonio Stella e Nikolaj Møller che hanno collaborato con entusiasmo a questo progetto. Un ringraziamento particolare va alla Prof.ssa e Presidentessa Stefania Serafin per il continuo sostegno dimostratomi. Questo studio è supportato da fondi del programma strategico 2016-2021 "Knowledge for the World" assegnati a MG da Aalborg University.

## 7. BIBLIOGRAFIA

- [1] M. Geronazzo, S. Spagnol, and F. Avanzini, "Mixed Structural Modeling of Head-Related Transfer Functions for Customized Binaural Audio Delivery," in *Proc. 18th Int. Conf. Digital Signal Process. (DSP 2013)*, (Santorini, Greece), pp. 1–8, July 2013.
- [2] M. Geronazzo, *Mixed structural models for 3D audio in virtual environments*. Ph.D. Thesis, University of Padova, Padova, Italy, Apr. 2014.
- [3] C. P. Brown and R. O. Duda, "A Structural Model for Binaural Sound Synthesis," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 6, no. 5, pp. 476–488, 1998.
- [4] J. Blauert, *The Technology of Binaural Listening*. Modern Acoustics and Signal Processing, Springer Berlin Heidelberg, 2013.
- [5] H. Gamper, *Enabling technologies for audio augmented reality systems*. PhD thesis, 2014.
- [6] J. Sandvad, "Dynamic Aspects of Auditory Virtual Environments," Audio Engineering Society, May 1996.
- [7] M. Geronazzo, J. Kleimola, and P. Majdak, "Personalization Support for Binaural Headphone Reproduction in Web Browsers," in *Proc. 1st Web Audio Conference*, (Paris, France), Jan. 2015.
- [8] S. Serafin, M. Geronazzo, N. C. Nilsson, C. Erkut, and R. Nordahl, "Sonic Interactions in Virtual Reality: State of the Art, Current Challenges and Future Directions," *IEEE Computer Graphics and Applications*, vol. 38, no. 2, pp. 31–43, 2018.
- [9] L. Savioja, J. Huopaniemi, T. Lokki, and R. Väänänen, "Creating interactive virtual acoustic environments," *Journal of the Audio Engineering Society*, vol. 47, no. 9, pp. 675–705, 1999.
- [10] B. B. Boren, M. Geronazzo, P. Majdak, and E. Choueiri, "PHOnA: A Public Dataset of Measured Headphone Transfer Functions," in *Proc. 137th Conv. Audio Eng. Society*, Audio Engineering Society, Oct. 2014.
- [11] H. Møller, M. Sørensen, J. Friis, B. Clemen, and D. Hammershøi, "Binaural Technique: Do We Need Individual Recordings?," *J. Audio Eng. Soc.*, vol. 44, no. 6, pp. 451–469, 1996.
- [12] M. Noisternig, B. F. Katz, S. Siltanen, and L. Savioja, "Framework for Real-Time Auralization in Architectural Acoustics," *Acta Acustica united with Acustica*, vol. 94, no. 6, pp. 1000–1015, 2008.
- [13] L. Picinali, A. Afonso, M. Denis, and B. F. G. Katz, "Exploration of architectural spaces by blind people using auditory virtual reality for the construction of spatial knowledge," *International Journal of Human-Computer Studies*, vol. 72, no. 4, pp. 393 – 407, 2014.
- [14] L. Savioja and U. P. Svensson, "Overview of geometrical room acoustic modeling techniques," *J. Acoust. Soc. Am.*, vol. 138, pp. 708–730, Aug. 2015.

- [15] R. Ranjan and W.-S. Gan, "Adaptive Equalization of Natural Augmented Reality Headset Using Non-Stationary Virtual Signals," 2016. 00000.
- [16] A. Deleforge, R. Horaud, Y. Schechner, and L. Girin, "Co-Localization of Audio Sources in Images Using Binaural Features and Locally-Linear Regression," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, pp. 718–731, Apr. 2015.
- [17] S. Braun, W. Zhou, and E. A. Habets, "Narrowband direction-of-arrival estimation for binaural hearing aids using relative transfer functions," in *2015 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pp. 1–5, Oct. 2015.
- [18] P. F. Hoffmann, A. K. Møller, F. Christensen, and D. Hammershøi, "Sound localization and speech identification in the frontal median plane with a hear-through headset," in *In Proc. 7th Forum Acusticum*, (Krakow, Poland), 2014. 00002.
- [19] M. Geronazzo, S. Spagnol, A. Bedin, and F. Avanzini, "Enhancing Vertical Localization with Image-guided Selection of Non-individual Head-Related Transfer Functions," in *IEEE Int. Conf. on Acoust. Speech Signal Process. (ICASSP 2014)*, (Florence, Italy), pp. 4496–4500, May 2014.
- [20] M. Geronazzo, S. Spagnol, and F. Avanzini, "Do We Need Individual Head-Related Transfer Functions for Vertical Localization? The Case Study of a Spectral Notch Distance Metric," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 26, pp. 1243–1256, July 2018.
- [21] B. Boren, M. Geronazzo, F. Brinkmann, and E. Choueiri, "Coloration Metrics for Headphone Equalization," in *Proc. of the 21st Int. Conf. on Auditory Display (ICAD 2015)*, (Graz, Austria), pp. 29–34, July 2015.
- [22] R. Albrecht, T. Lokki, and L. Savioja, "A mobile augmented reality audio system with binaural microphones," in *Proceedings of Interacting with Sound Workshop: Exploring Context-Aware, Local and Social Audio Applications*, IwS '11, (Stockholm, Sweden), pp. 7–11, ACM, 2011.
- [23] M. Geronazzo, J. Fantin, G. Sorato, G. Baldovino, and F. Avanzini, "Acoustic Selfies for Extraction of External Ear Features in Mobile Audio Augmented Reality," in *Proc. 22nd ACM Symposium on Virtual Reality Software and Technology (VRST 2016)*, (Munich, Germany), pp. 23–26, ACM, Nov. 2016.
- [24] A. Deleforge, F. Forbes, and R. Horaud, "Acoustic space learning for sound-source separation and localization on binaural manifolds," *International Journal of Neural Systems*, vol. 25, no. 01, p. 1440003, 2015.
- [25] M. Geronazzo, J. Kleimola, E. Sikström, A. de Götzen, S. Serafin, and F. Avanzini, "HOBA-VR: HRTF On Demand for Binaural Audio in Immersive Virtual Reality Environments," (Milano), Audio Engineering Society, May 2018.
- [26] M. Geronazzo, A. Carraro, and F. Avanzini, "Evaluating vertical localization performance of 3d sound rendering models with a perceptual metric," in *2015 IEEE 2nd VR Workshop on Sonic Interactions for Virtual Environments (SIVE)*, (Arles, France), pp. 1–5, IEEE Computer Society, Mar. 2015.
- [27] R. Barumerli, M. Geronazzo, and F. Avanzini, "Localization in Elevation with Non-Individual Head-related Transfer Functions: Comparing Predictions of Two Auditory Models," in *Proc. 26th European Sig. Proc. Conf. (EUSIPCO 2018) - accepted for publication*, (Rome, Italy), IEEE Signal Processing Society, Sept. 2018.
- [28] A. Lindau, V. Erbes, S. Lepa, H.-J. Maempel, F. Brinkman, and S. Weinzierl, "A Spatial Audio Quality Inventory (SAQI)," *Acta Acustica united with Acustica*, vol. 100, pp. 984–994, Sept. 2014.
- [29] L. S. R. Simon, N. Zacharov, and B. F. G. Katz, "Perceptual attributes for the comparison of head-related transfer functions," *The Journal of the Acoustical Society of America*, vol. 140, pp. 3623–3632, Nov. 2016.
- [30] M. Geronazzo, E. Peruch, F. Prandoni, and F. Avanzini, "Improving elevation perception with a tool for image-guided head-related transfer function selection," in *Proc. of the 20th Int. Conference on Digital Audio Effects (DAFx-17)*, (Edinburgh, UK), pp. 397–404, Sept. 2017.
- [31] H. Bahu and D. Romblom, "Optimization and Prediction of the Spherical and Ellipsoidal ITD Model Parameters Using Offset Ears," Audio Engineering Society, July 2018.
- [32] E. De Sena, H. Hacıhabiboglu, Z. Cvetkovic, and J. O. Smith, "Efficient Synthesis of Room Acoustics via Scattering Delay Networks," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, pp. 1478–1492, Sept. 2015.
- [33] L. McCormack, S. Delikaris-Manias, and V. Pulkki, "Parametric acoustic camera for real-time sound capture, analysis and tracking," p. 8, 2017.
- [34] R. Albrecht, J. Jaatinen, and T. Lokki, "Electronic hearing protection for musicians," in *Sound and Music Computing Conference*, 2017.
- [35] J. Oberem, I. Koch, and J. Fels, "Intentional switching in auditory selective attention: Exploring age-related effects in a spatial setup requiring speech perception," *Acta Psychologica*, vol. 177, pp. 36–43, 2017.
- [36] R. F. Lyon, *Human and Machine Hearing: Extracting Meaning from Sound*. Cambridge: Cambridge University Press, 2017.

# Lightweight Virtual Analog Modeling

**Stefano D'Angelo**  
 Independent Researcher  
 Via P. Voso 22/B  
 Agropoli 84043, Italy  
 s@dangelo.audio

## ABSTRACT

The purpose of this paper is to give theoretical and practical advice to DSP programmers willing to implement efficient real-time circuit simulation algorithms. It summarizes the fundamental concepts behind the most widespread circuit modeling techniques and gives the reader sufficient references to get started. Then it tackles on a certain number of topics that are often neglected in theoretical studies yet highly relevant in real-world applications. Hopefully, this paper might also serve as a reference and an invite for researchers to direct their efforts towards problems and solutions that have a concrete impact.

## 1. INTRODUCTION

Many audio processing algorithms developed as of today are to some degree inspired by or try to replicate real-world phenomena [1, 2, 3], whether directly sound-related or where sound manipulation happens as a byproduct of another process. In particular, modeling of analog circuits that generate or process sound-related signals sparked particular interest in the last decades, both in industry and academia [4, 5, 6].

Simulation or emulation of circuits for real-time sound-related applications is often called “Virtual Analog modeling” (VA modeling), and many approaches have been proposed to address it specifically [7, 8]. Obviously, each technique has its peculiar strengths and weaknesses, yet some specific and often neglected aspects tend to be of utmost importance in practical applications: computational cost, maintainability, and parameterization.

This paper very concisely reviews the most widespread white-box VA modeling frameworks in Section 2 and provides some practical modeling and implementation advice in Section 3. Section 4 concludes the paper.

Copyright: ©2018 Stefano D'Angelo et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution License 3.0 Unported](https://creativecommons.org/licenses/by/3.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

## 2. VA MODELING

### 2.1 State-space methods

State-space (SS) representation can be used to describe analog electronic circuits [9, 10, 11, 12] as

$$\frac{dx}{dt} = \mathbf{A}x + \mathbf{B}u + \mathbf{C}i, \quad (1)$$

$$\mathbf{i} = f(\mathbf{v}), \quad (2)$$

$$\mathbf{v} = \mathbf{D}x + \mathbf{E}u + \mathbf{F}i, \quad (3)$$

$$\mathbf{y} = \mathbf{L}x + \mathbf{M}u + \mathbf{N}i, \quad (4)$$

where  $\mathbf{x}$  is the state vector,  $\mathbf{u}$  is the input vector,  $\mathbf{y}$  is the output vector, and  $f(\cdot)$  is a MIMO nonlinear mapping,  $t$  is the time variable, and  $\mathbf{A}$  to  $\mathbf{N}$  are matrices describing the linear behavior of the system.

Such a system can be discretized by means of common numerical integration schemes, such as the backward Euler method or the trapezoidal rule, and finally implemented, usually after further mathematical manipulation. In most cases, however, implicit nonlinear equations arise, which need to be solved numerically or otherwise approximated.

### 2.2 Wave digital filters

Wave digital filters (WDFs) represent another technique to model and discretize analog circuits [13, 14, 15, 16] which does not directly operate on Kirchoff variables (*i.e.*, voltages and currents), but rather on so-called *wave variables* defined as

$$\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 1 & R_0 \\ 1 & -R_0 \end{pmatrix} \begin{pmatrix} V \\ I \end{pmatrix}, \quad (5)$$

where  $a$  is the incoming wave,  $b$  is the reflected wave,  $R_0$  is a free parameter, and  $V$  and  $I$  are, respectively, the voltage and current of an electrical port. By this formulation, all LTI multiports (*e.g.*, resistors, capacitors, inductors, terminated voltage and current sources) can be translated into standalone and reusable DSP components that can be interconnected using a tree of *adaptors* implementing the topology of the circuit.

It is remarkable that a WDF tree represents, at the same time, the model and the DSP structure of the resulting algorithm. On the other hand, in their classic formulation, WDFs are only suitable for a limited set of topologies and up to one nonlinear element. Several solutions have been proposed to overcome such limitations [17, 18, 19, 20].

### 2.3 Port-Hamiltonian systems

An analog circuit can also be modeled as a port-Hamiltonian system (PHS) [21, 22, 23] by the equation

$$\begin{pmatrix} \frac{dx}{dt} \\ \mathbf{w} \\ \mathbf{y} \end{pmatrix} = \begin{pmatrix} \mathbf{J}_x & -\mathbf{K} & \mathbf{G}_x \\ \mathbf{K}^T & \mathbf{J}_w & \mathbf{G}_w \\ \mathbf{G}_x^T & \mathbf{G}_w^T & \mathbf{J}_y \end{pmatrix} \begin{pmatrix} \nabla \mathcal{H}(\mathbf{x}) \\ \mathbf{z}(\mathbf{w}) \\ \mathbf{u} \end{pmatrix}. \quad (6)$$

PHS formulation is a subclass of SS representation with a special focus on energy balancing, and therefore naturally tends to the development of algorithms with remarkable stability. The downsides are, arguably, a higher level of theoretical complexity and higher resulting computational cost when compared to other options.

## 3. PRACTICAL CONSIDERATIONS

When developing digital models of analog circuits for the general public, the mechanical application of rules found in scientific literature rarely does directly lead to acceptable results. This section explores a few practical aspects of VA modeling.

### 3.1 Programming approach

Computational cost is a major concern in real-time systems, therefore DSP code is often written in languages such as C and C++ that are nowadays considered relatively low-level and for which highly-optimizing compilers are available, or even in machine-specific assembly. These languages, however, are based on imperative programming paradigms that are quite distant from the dataflow metaphor, which more naturally describes most DSP applications [24, 25]. Unsurprisingly, there have been some attempts at developing domain-specific languages for the task, and especially based off declarative syntaxes and coupled with a source-to-source compiler that emits C or C++ code [26, 27, 28, 29].

To the best of my knowledge, however, none of them has been widely adopted by the industry yet, and most existing codebases remain written in traditional general-purpose languages. Therefore, for the sake of code correctness and maintainability, I would recommend DSP programmers to rather concentrate on mathematical aspects, DSP structure design, and prototyping, and to defer the actual coding to the latest possible development stages. Furthermore, I would suggest to first translate prototypes and designs to production code in a very mechanical fashion, that is following clearly-defined patterns and avoiding language-provided abstractions, and only apply optimizations at the very end and only where needed. In the end, it is also highly recommendable to produce clear and exhaustive documentation and to keep it up to date.

### 3.2 Mathematical modeling

The derivation of workable models of analog circuits is a repetitive, tedious, and error-prone activity, which can luckily be largely automated. Traditionally, the go-to method for expressing a circuit in SS representation is Modified Nodal Analysis (MNA) [30], which can be found at the

heart of circuit simulators such as SPICE [31], and which still influences recent automatic modeling approaches and tools [10, 11, 32]. Also PHSs can be automatically formulated from schematics [23], and indeed a tool called PyPHS [33] exists that goes as far as generating C++ code that simulates a given circuit. Finally, tools to ease the implementation of WDFs have also been proposed [26, 27, 29, 34].

The manipulation of model equations can easily become cumbersome to perform manually, hence the usage of mathematical software is nowadays indispensable. At the very least, one should have access to software for numerical analysis, symbolic computation, arbitrary-precision arithmetic, 2D/3D plotting, and a prototyping language.

Models should always be verified by comparison with a reliable circuit simulator, such as SPICE, and possibly with measurements of real-world devices. In this sense, while it is necessary to choose appropriate input test signals on a case-by-case basis, it is still also possible to employ well-documented output analysis techniques [35, 36]. In the end, it is also advisable to refrain from the temptation of excessively complicating a model in order to match measurements, as certain improvements might fall under the thresholds of component tolerances or human hearing.

### 3.3 Circuit separation

The computational complexity of digital models of nonlinear analog circuits tends to grow more-than-linearly with the number of components involved, hence large savings can be obtained by simply splitting a circuit into two or more subcircuits characterized by unidirectional control flow [8, 37, 38]. In most cases, this corresponds to an approximation that may or may not be acceptable for a given application, therefore careful evaluation is needed. The definition of an automated method for the task is still an open research topic [39]. Furthermore, an alternative approach that has been proposed is the joint simulation of adjacent couples of nonlinear subcircuits in a chain [40, 38].

### 3.4 Delay-free loop implementation

Another major source of computational issues is the presence of implicit nonlinear equations, which naturally translate to delay-free loops in DSP terms. The numerical solution of such equations through root-finding methods is a central topic of numerical analysis, and indeed it is rather common to employ such iterative methods [41, 42, 43, 22].

Otherwise, it is possible to employ precomputed lookup tables [9, 19]. This approach is generally only viable as long as the dimensionality of such tables is low and the input ranges are known beforehand. Also, the use of tables may substantially increase cache pressure, thus potentially causing significant and hard-to-detect performance problems.

In the case of simple transcendental equations, which are often obtained when modeling circuits containing semiconductor devices, closed-form solutions or remarkably accurate approximations have been found [44, 45, 46, 47]. Furthermore, piecewise-linear models have also been proposed to obtain approximate solutions of delay-free loops [48, 49].

In some cases the problem was avoided by simply adding fictitious delay units in the feedback branches of delay-free loops [50, 51]. While extremely simple and efficient, usage of this method affects the frequency response of the system and often also leads to undesirable and nonobvious parameter coupling effects. If a good amount of oversampling is used, the outcome can actually be acceptable, otherwise a method has also been proposed to restore the original linear behavior and parameterization by adding extra compensation filters [52]. The stability and accuracy properties of such method, however, still await systematic study.

### 3.5 Further optimizations

Oversampling can be yet another source of inefficiency. Firstly, recent research has gone into alias reduction without oversampling [53, 54]. Then, it is fundamental to carefully choose oversampling ratios and resampling boundaries when using chains of oversampled structures. Finally, in many cases IIR resampling filters provide acceptable results and use less resources than FIR filters.

Beforehand knowledge on, e.g., signal and parameter ranges and frequency-related effects, can justify convenient approximations and substantial simplifications without affecting the output quality significantly. The computation of nonlinear functions can be sped up significantly by employing polynomial and rational approximations, leveraging binary number representations, or avoiding general-purpose routines. Also, limiting the computation of coefficients only to the moment in which a parameter value is changed and precomputing as many subexpressions as possible are strategies that allow for significant savings. Finally, while modern compilers are said to perform excellent code optimization, in my experience it is still possible to exploit hardware-specific knowledge and facilities, such as vector instructions, in order to reduce computational load even more.

## 4. CONCLUSIONS

The translation of analog circuits into actual production-ready VA models is a highly multidisciplinary activity. It requires knowledge not only regarding the specific topic, but also in mathematics, circuit analysis, computer science, DSP, psychoacoustics, as well as familiarity with specific software tools and a decent amount of self-discipline. The development of lightweight and maintainable VA models is possible by taking into account many different aspects, especially w.r.t. optimizations and approximations, and always being ready to explore non-systematic approaches. In other words, as of today, high-performance VA modeling is still be more of a craft than a predetermined and automatable industrial task.

## 5. REFERENCES

- [1] J. O. Smith, "Physical modeling synthesis update," *Computer Music J.*, vol. 20, no. 2, pp. 44–56, 1996.
- [2] V. Välimäki, J. Pakarinen, C. Erkut, and M. Karjalainen, "Discrete-time modelling of musical instruments," *Reports on Progress in Physics*, vol. 69, pp. 1–78, January 2006.
- [3] V. Välimäki, J. D. Parker, L. Savioja, J. O. Smith, and J. S. Abel, "More than 50 years of artificial reverberation," in *Proc. AES 60<sup>th</sup> Conf.: DREAMS (Dereverberation and Reverberation of Audio, Music, and Speech)*, (Leuven, Belgium), February 2016.
- [4] V. Välimäki, F. Fontana, J. O. Smith, and U. Zölzer, "Introduction to the special issue on virtual analog audio effects and musical instruments," *IEEE Trans. Audio, Speech, and Lang. Process.*, vol. 18, pp. 713–714, May 2010.
- [5] V. Välimäki, S. Bilbao, J. O. Smith, J. S. Abel, J. Pakarinen, and D. P. Berners, "Virtual analog effects," in *DAFX: Digital Audio Effects, Second Edition* (U. Zölzer, ed.), pp. 473–522, Chichester, UK: Wiley, 2011.
- [6] J. Pakarinen, V. Välimäki, F. Fontana, V. Lazzarini, and J. S. Abel, "Recent advances in real-time musical effects, synthesis, and virtual analog models," *EURASIP J. Advances Signal Process.*, vol. 2011, Article ID 940784, 15 pages, 2011.
- [7] R. C. D. de Paiva, *Circuit modeling studies related to guitars and audio processing*. PhD thesis, Aalto University, Espoo, Finland, November 2013.
- [8] S. D'Angelo, *Virtual analog modeling of nonlinear musical circuits*. PhD thesis, Aalto University, Espoo, Finland, November 2014.
- [9] G. Borin, G. De Poli, and D. Rocchesso, "Elimination of delay-free loops in discrete-time models of nonlinear acoustic systems," *IEEE Trans. Audio, Speech, and Lang. Process.*, vol. 8, pp. 597–605, September 2000.
- [10] D. Yeh, J. Abel, and J. Smith, "Automated physical modeling of nonlinear audio circuits for real-time audio effects—part I: Theoretical development," *IEEE Trans. Audio, Speech, and Lang. Process.*, vol. 18, pp. 728–737, May 2010.
- [11] D. Yeh, "Automated physical modeling of nonlinear audio circuits for real-time audio effects—part II: BJT and vacuum tube examples," *IEEE Trans. Audio, Speech, and Lang. Process.*, vol. 20, pp. 1207–1216, May 2012.
- [12] K. Dempwolf, M. Holters, and U. Zölzer, "Discretization of parametric analog circuits for real-time simulations," in *Proc. 13<sup>th</sup> Intl. Conf. Digital Audio Effects (DAFx-10)*, (Graz, Austria), September 2010.
- [13] A. Fettweis, "Wave digital filters: Theory and practice," *Proc. IEEE*, vol. 74, pp. 270–327, February 1986.
- [14] S. Bilbao, *Wave and scattering methods for numerical simulation*. John Wiley & Sons, 2004.

- [15] G. De Sanctis and A. Sarti, “Virtual analog modeling in the wave-digital domain,” *IEEE Trans. Audio, Speech, and Lang. Process.*, vol. 18, pp. 715–727, May 2010.
- [16] K. J. Werner, *Virtual analog modeling of audio circuitry using wave digital filters*. PhD thesis, Stanford University, Stanford, USA, December 2016.
- [17] S. Petrausch and R. Rabenstein, “Wave digital filters with multiple nonlinearities,” in *Proc. 12<sup>th</sup> European Signal Process. Conf. (EUSIPCO 2004)*, (Wien, Austria), pp. 77–80, September 2004.
- [18] D. Fränken, J. Ochs, and K. Ochs, “Generation of wave digital structures for networks containing multiport elements,” *IEEE Trans. Circ. Systems I: Regular Papers*, vol. 52, pp. 586–596, March 2005.
- [19] K. J. Werner, V. Nangia, J. O. Smith, and J. S. Abel, “Resolving wave digital filter with multiple/multiport nonlinearities,” in *Proc. 18<sup>th</sup> Intl. Conf. Digital Audio Effects (DAFx-15)*, (Trondheim, Norway), November 2015.
- [20] K. J. Werner, J. O. Smith, and J. S. Abel, “Wave digital filter adaptors for arbitrary topologies and multiport linear elements,” in *Proc. 18<sup>th</sup> Intl. Conf. Digital Audio Effects (DAFx-15)*, (Trondheim, Norway), November 2015.
- [21] A. Falaize and T. Hélie, “Simulation of an analog circuit of a wah pedal: A port-Hamiltonian approach,” in *Proc. 135<sup>th</sup> AES Convention*, (New York, USA), October 2013.
- [22] N. Lopes, T. Hélie, and A. Falaize, “Explicit second-order accurate method for the passive guaranteed simulation of port-Hamiltonian systems,” *IFAC-PapersOnLine*, vol. 48, no. 13, pp. 223–228, 2015.
- [23] A. Falaize and T. Hélie, “Passive guaranteed simulation of analog audio circuits: A port-Hamiltonian approach,” *Applied Sciences*, vol. 6, no. 10, 2016.
- [24] E. A. Lee and D. G. Messerschmitt, “Synchronous data flow,” *Proc. IEEE*, vol. 75, pp. 1235–1245, September 1987.
- [25] K. Barkati and P. Jouvelot, “Synchronous programming in audio processing: A lookup table oscillator case study,” *ACM Computing Surveys (CSUR)*, vol. 46, pp. 24:1–24:35, December 2013.
- [26] M. Karjalainen, “BlockCompiler: Efficient simulation of acoustic and audio systems,” in *Proc. 114<sup>th</sup> AES Convention*, (Amsterdam, Netherlands), March 2003.
- [27] M. Karjalainen, “BlockCompiler - a research tool for physical modeling and DSP,” in *Proc. 6<sup>th</sup> Intl. Conf. Digital Audio Effects (DAFx-03)*, (London, UK), September 2003.
- [28] Y. Orlarey, D. Fober, and S. Letz, “FAUST: An efficient functional approach to DSP programming,” *New Computational Paradigms for Computer Music*, vol. 290, 2009.
- [29] S. D’Angelo, “Physics-based modeling for sound processing: Theory and implementation,” Master’s thesis, Politecnico di Torino, Turin, Italy, September 2010.
- [30] C.-W. Ho, A. E. Ruehli, and P. A. Brennan, “The modified nodal approach to network analysis,” *IEEE Trans. Circ. Systems*, vol. 22, pp. 504–509, June 1975.
- [31] L. W. Nagel, *SPICE2: A computer program to simulate semiconductor circuits*. PhD thesis, University of California, Berkeley, USA, May 1975.
- [32] M. Holters and U. Zölzer, “A generalized method for the derivation of non-linear state-space models from circuit schematics,” in *Proc. 23<sup>rd</sup> European Signal Process. Conf. (EUSIPCO 2015)*, (Nice, France), pp. 1073–1077, IEEE, August 2015.
- [33] “PyPHS.” <http://pyphs.github.io/pyphs/>. Accessed September 16<sup>th</sup>, 2018.
- [34] M. Rest, R. W. Dunkel, K. J. Werner, and J. O. Smith, “RT-WDF—a modular wave digital filter library with support for arbitrary topologies and multiple nonlinearities,” in *Proc. 19<sup>th</sup> Intl. Conf. Digital Audio Effects (DAFx-16)*, (Brno, Czech Republic), September 2016.
- [35] A. Farina, “Simultaneous measurement of impulse response and distortion with a swept-sine technique,” in *Proc. 108<sup>th</sup> AES Convention*, (Paris, France), February 2000.
- [36] J. Pakarinen, “Distortion analysis toolkit—a software tool for easy analysis of nonlinear audio systems,” *EURASIP J. Advances Signal Process.*, vol. 2010, Article ID 617325, 13 pages, 2010.
- [37] O. Bogason and K. J. Werner, “Modeling circuits with operational transconductance amplifiers using wave digital filters,” in *Proc. 20<sup>th</sup> Intl. Conf. Digital Audio Effects (DAFx-17)*, (Edinburgh, UK), September 2017.
- [38] J. Zhang and J. O. Smith, “Real-time wave digital simulation of cascaded vacuum tube amplifiers using modified blockwise method,” in *Proc. 21<sup>th</sup> Intl. Conf. Digital Audio Effects (DAFx-18)*, (Aveiro, Portugal), September 2018.
- [39] M. Holters and U. Zölzer, “Automatic decomposition of non-linear equation systems in audio effect circuit simulation,” in *Proc. 20<sup>th</sup> Intl. Conf. Digital Audio Effects (DAFx-17)*, (Edinburgh, UK), September 2017.
- [40] J. Mačák, *Real-time digital simulation of guitar amplifiers as audio effects*. PhD thesis, Brno University of Technology, Brno, Czech Republic, 2012.

- [41] F. Fontana, F. Avanzini, and D. Rocchesso, "Computation of nonlinear filter networks containing delay-free paths," in *Proc. 7<sup>th</sup> Intl. Conf. Digital Audio Effects (DAFx-04)*, (Naples, Italy), pp. 113–118, October 2004.
- [42] F. Fontana and E. Bozzo, "Explicit fixed-point computation of nonlinear delay-free loop filter networks," *IEEE/ACM Trans. Audio, Speech, and Lang. Process.*, vol. 26, pp. 1884–1896, October 2018.
- [43] M. J. Olsen, K. J. Werner, and J. O. Smith, "Resolving grouped nonlinearities in wave digital filters using iterative techniques," in *Proc. 19<sup>th</sup> Intl. Conf. Digital Audio Effects (DAFx-16)*, (Brno, Czech Republic), September 2016.
- [44] J. Parker, "A simple digital model of the diode-based ring-modulator," in *Proc. 14<sup>th</sup> Intl. Conf. Digital Audio Effects (DAFx-11)*, (Paris, France), September 2011.
- [45] R. C. D. de Paiva, S. D'Angelo, J. Pakarinen, and V. Välimäki, "Generation of wave digital structures for networks containing multiport elements," *IEEE Trans. Circ. Systems II: Express Briefs*, vol. 59, pp. 688–692, October 2012.
- [46] J. Parker and S. D'Angelo, "A digital model of the Buchla lowpass-gate," in *Proc. 16<sup>th</sup> Intl. Conf. Digital Audio Effects (DAFx-13)*, (Maynooth, Ireland), September 2013.
- [47] A. Bernardini, K. J. Werner, A. Sarti, and J. O. Smith, "Modeling nonlinear wave digital elements using the Lambert function," *IEEE Trans. Circ. Systems I: Regular Papers*, vol. 63, pp. 1231–1242, August 2016.
- [48] A. Bernardini, K. J. Werner, A. Sarti, and J. O. Smith, "Modeling a class of multi-port nonlinearities in wave digital structures," in *Proc. 23<sup>rd</sup> European Signal Process. Conf. (EUSIPCO 2015)*, (Nice, France), pp. 664–668, August 2015.
- [49] A. Bernardini and A. Sarti, "Canonical piecewise-linear representation of curves in the wave digital domain," in *Proc. 25<sup>th</sup> European Signal Process. Conf. (EUSIPCO 2017)*, (Kos, Greece), pp. 1125–1129, August 2017.
- [50] A. Huovilainen, "Non-linear digital implementation of the Moog ladder filter," in *Proc. 7<sup>th</sup> Intl. Conf. Digital Audio Effects (DAFx-04)*, (Naples, Italy), pp. 61–64, October 2004.
- [51] M. Karjalainen and J. Pakarinen, "Wave digital simulation of a vacuum-tube amplifier," in *Proc. Intl. Conf. Acoust., Speech, and Signal Process. (ICASSP 2006)*, vol. 5, (Toulouse, France), pp. 153–156, May 2006.
- [52] S. D'Angelo and V. Välimäki, "Generalized moog ladder filter: Part iexplicit nonlinear model through a novel delay-free loop implementation method," *IEEE/ACM Trans. Audio, Speech, and Lang. Process.*, vol. 22, pp. 1873–1883, December 2014.
- [53] F. Esqueda, *Aliasing reduction in nonlinear audio signal processing*. PhD thesis, Aalto University, Espoo, Finland, May 2018.
- [54] E. Paschou, F. Esqueda, V. Välimäki, and J. Mourjopoulos, "Modeling and measuring a Moog voltage-controlled filter," in *2017 Asia-Pacific Signal and Information Process. Association Annual Summit and Conference (APSIPA ASC)*, pp. 1641–1647, December 2017.

# *Rima Glottidis*: Experimenting Generative Raw Audio Synthesis for a Sound Installation

Leonardo Gabrielli, Fabio Vesperini, Diego Droghini, Stefano Squartini  
 Università Politecnica delle Marche, Ancona, Italy

## ABSTRACT

Biologically-inspired algorithms such as artificial neural networks have been used extensively by computer music researchers, for generative music and algorithmic composition. The recent introduction of raw audio Machine Learning (ML) techniques, however, represents a significant leap because they seem to be able to learn both high-level (event) and low-level (timbre) information at once. Employing such techniques for creative purposes is very challenging at this early stage since there is lack of method, experience, and their computational cost is very high. In this paper, we describe the technical and creative process behind *Rima Glottidis*, an installation based on material from the homonym musical work from the artist *Okapi*, processed using raw audio ML techniques for an installation premiered at Blooming festival, Italy. Technical issues are described and critical aspects are reported as a reference for future projects.

## 1. INTRODUCTION

The computer music research community has always been looking with interest towards biologically-inspired algorithms [1] such as genetic algorithms [2, 3], cellular automata [4, 5], artificial neural networks (ANN) [3, 6, 7] and other nonlinear dynamical systems [8, 9]. All these algorithms have been used mostly for generative music and algorithmic composition, with some exceptions (see e.g. the works in [8, 10, 11]). In the past years, ANN were mainly suited to offline operation and could hardly manage raw signals. However, the recent introduction of end-to-end machine learning techniques, and, more specifically, raw audio architectures, represents a significant leap in the semantic handling of musical signals because these seems to be able to represent high-level information such as events on a human scale (notes, phonemes, etc.) and low-level information, (i.e. timbre) at once. Handling these techniques is, nonetheless, very complex and computationally expensive.

To the best of our knowledge, three main architectures have been proposed to learn and reproduce features from raw audio signals, namely WaveNet [12], sampleRNN [13],

and WaveGan [14]. The WaveNet structure has also been coupled with its inverse for creating a neural autoencoder [15]. Recently, another architecture, named FFTnet, has been proposed [16] that draws some concepts from WaveNet but shows a lower computational cost.

These architectures show the ability of learning features from raw audio data and generate similar signals drawing from a *latent* representation of these data directly in the time domain. This seems to be possible thanks to a hierarchical representation of the signal. In other words, architectures such as sampleRNN and WaveNet process the signal at multiple levels, enabling the network to store significant features from the micro-scale (sample level) up to the macro-scale (1 second and more). Experiments with these two architectures show that after training the network with a few hours of piano music, the networks are able to reproduce piano tones in a somewhat organized way. They are, thus, able to learn the timbre of the piano, the piano tone envelope (attack and decay), and finally, they learn that piano notes have rhythmical patterns, can be played in clusters, chords and sequences.

Stemming from these results, the authors decided to explore the possibility of running these algorithms for tone generation. At the time of writing, these algorithms are not suitable for synthesis from a score. However, they have the autonomous ability of generating a sequence of sounds, similarly to generative algorithms, but working at several time scales.

In this work we describe the design of a sound installation based on these algorithms and we report the outcomes of the experiments, without the intent of being exhaustive. The work is meant as an exploration and will hopefully suggest ideas to the computer music community and its artists to develop further contents.

## 2. PROJECT WORKFLOW

Motivation for this work comes from digital arts festival Blooming<sup>1</sup> commissioning an arts installation as contribution to the festival programme. The only constraints for this installation were the exploitation of novel digital technologies and the location where the installation would take place, an ancient cellar, i.e. an artificial grotto, excavated under a noble palace of the town of Pergola, Italy, employed during the centuries for defensive and civil uses.

Artistic supervision of the work was curated by the Italian cut-up artist *Okapi*, who adheres to the plunderphonics

Copyright: ©2018 Leonardo Gabrielli, Fabio Vesperini, Diego Droghini, Stefano Squartini et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution License 3.0 Unported](https://creativecommons.org/licenses/by/3.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

<sup>1</sup> [www.bloomingfestival.it](http://www.bloomingfestival.it)



**Figure 1:** A picture of the ancient cellar where the installation took place.

aesthetic and collects audio material from a lot of different sources. It was settled that he would provide the input material to train neural networks.

## 2.1 Algorithm Selection

A careful evaluation of currently available machine learning algorithms for raw-audio generation has been carried on. The evaluated algorithms were:

- WaveNet [12],
- sampleRNN [13],
- WaveGAN [14].

At the time of setting up the experiments the FFTNet algorithm [16] was not yet published.

The WaveNet architecture allows for modeling complex data such as music and speech. It is based on a stack of causal dilated convolutional layers for feature extraction. The dilated convolutions are key to extract features at different levels (closely resembling a dyadic wavelet filter bank [16]), however, the experiments in [17] show that the use of dilated convolutions alone do not allow modelling complex raw audio sequences, thus, residual blocks and the use of skip connections after the convolutional layers are necessary to speed up convergence and allow the training of a deep model. The ability of the network to model meaningful sequences of samples relies on causal probability conditioning, i.e. the last output sample is conditioned by the probability of all previous output samples. Additionally, it has been shown with speech synthesis that the output sample probability can be conditioned with, e.g. speaker timbre and phonemes. In our case, however, we are not interested in leveraging this feature, leaving the network to generate freely.

The main issue preventing adoption of WaveNet is its high computational cost and informal reports of their performance seems to confirm this. To test the feasibility of the approach and verify these claims we adapted one of the several open-source implementations of WaveNet currently available. We conducted preliminary experiments on a Titan X GPU employing Theano as backend. After 72

hours of training using the same piano dataset suggested by [12] the network was still unable to provide results vaguely similar to the ones shown by the authors of [12]. Furthermore, some of the hyperparameters were extrapolated from the paper, but most of them are not available, so a hyperparameters search would be required, increasing the training times far over our available resources.

A very different approach is followed by the authors of sampleRNN [13]. This architecture is based on the same principle of causal probability conditioning, however, in this case the network consists of several tiers of recurrent neural networks (RNN) working at different temporal resolutions. Each RNN conditions the RNN at the lower tier. SampleRNN has been tested with both speech and music database. Both database are a few hours long.

We also performed preliminary experiments employing the open-source implementation provided by the authors of [13] on the same hardware reported above. This implementation is based on Theano and Lasagne. Results are on par with those provided by the authors and training times can have length of 1-4 days depending on the input material and the degree of fitness that is required.

Finally, we considered WaveGAN for sound generation. This architecture is based on the generative adversarial networks (GAN) paradigm, where two neural networks, a generator  $G$  and a discriminator  $D$ , are employed for training. The role of  $G$  is to generate plausible outputs (e.g. images), while the role of  $D$  is to discriminate whether an example is real or is generated by  $G$ . One is, thus, trained to minimize a cost function, while the other one is trained to maximize it. The authors of [14] devised two models, WaveGAN, working on the raw audio with 1D convolutions and SpecGAN, working with 2D spectrograms. The outcomes of their work are very promising, since these two architecture are able to learn from a much shorter database than WaveNet and sampleRNN. However, they are inherently designed to synthesized audio of a specific length (16384 samples for WaveGAN or a 128x128 spectrogram for SpecGAN). For this reason they are not well suited for generative audio. Running a WaveGAN several times would introduce issues in concatenating each output.

At the end of this evaluation stage, we decided to use sampleRNN. Some additional material has been generated using a recent neural timbre transfer algorithm proposed by the authors in [18], leaving more freedom for the arrangement stage.

## 2.2 Input Material and Dataset Creation

The input material for this work has been provided by artist *Okapi* in form of *stems* of his musical project Rima Glottidis. The main goal of the work was to generate vocal textures from his material to be arranged in form of a site-specific sound installation. The stems came from cut recordings of male and female speech of several speakers showing proper Italian pronunciation vs. unclear spelling and regional accent.

Additionally they presented bell tones, singing choirs, sung phonemes, and silence, revealing part of the musical

structure of the original project. The whole material length was 14 minutes. A subset of this material, totalling 7 minutes, was obtained by leaving only the speaking voices and cutting all other musical material.

From previous experience with sampleRNN the input material was judged to be too short. Informal experiments on voice synthesis were conducted in 2016 by the first author showing that sampleRNN can faithfully learn the vocal timbre of a speaker only from a sufficiently large and homogeneous dataset of speech. Specifically, a 3 hours dataset gathered from publicly available speeches of a well known Italian comedian and politician was used to train a three-tier sampleRNN model. At the end of the training, a convincing babbling was obtained with random phonemes. Most subjects presented with the synthesized speech could recognize the identity of the public figure. Reducing the dimension of the dataset to 1 hour or less degraded the performance of the network making the utterances noise-like.

To increase the chances to obtain interesting audio material from sampleRNN, other audio material has been selected. Data augmentation has been tested to enlarge the training corpus without adding spurious material that would reduce the portion of material from *Økapi*. Data augmentation has been done by pitch shifting [19] audio data by -3 and +3 semitones.

Additionally, 22 minutes of choral works from Arvo Pärt were added. A subset of this dataset has been randomly selected, totalling 2 minutes.

In conclusion several datasets have been composed from the above material:

- Rima Glottidis (14 min)
- Rima Glottidis speech-only subset (7 min)
- Rima Glottidis and pitch shifted versions (42 min)
- Rima Glottidis subset and pitch shifted versions (21 min)
- Rima Glottidis and Chorales (36 min)
- Rima Glottidis and Chorales subset (16 min)
- Rima Glottidis and Chorales, all pitch shifted (108 min)
- Chorales (22 min)

Each of these datasets have been used for training 2 different sampleRNN models with hyperparameters suggested by the authors of [13]. Training of these models was conducted with a Titan X GPU and left running for 2 days for each one, approximately lasting for 220 epochs (approx  $3 \cdot 10^6$  iterations). For each epoch, several 10s-long pieces of randomly generated material were extracted for later use.

### 2.3 Outcomes

Discussion about the network outcomes will be qualitative, as there is no objective means to assess the quality of the material. The loss employed by sampleRNN, the categorical crossentropy, does not clearly state the quality of the

sound material. The largest excursion of the loss seen during training ranged from 5 to 3 for the training set and from 4.5 to 3.9 for the validation set. Notwithstanding this, the first tones generated were pure noise, while the last ones had some of the character of the original dataset. The training loss did not necessarily decrease with time and samples with interesting features emerged at different epochs. Most notably, samples generated at a given epoch or at consecutive epochs, have all very similar features that later disappear with the training.

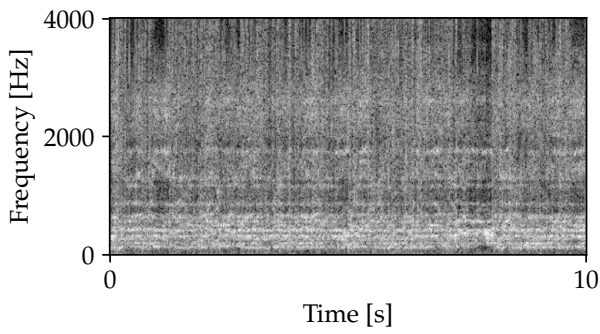
Among the datasets employed, the one generating most predictable results was the Chorales one. With this dataset after only 4 epochs a choir texture emerges and remains, with slight variations present for all the training, with the only difference that, with the epochs passing, the texture have more swift shifts in the formants. The material obtained from this dataset would be very interesting for use in such a context. Unfortunately, there is a lot of signal clipping. This is partly due to the low dynamic range of the network, which outputs 8-bit values, even though A-law companding is employed to improve the dynamic range. There are known examples of non-distorted audio generated by sampleRNN, suggesting that there might be ways to mitigate this distortion with further experimentation. Samples for Rima Glottidis were taken after cutting away the distorted parts or by reducing clipping.

Results with all the other datasets are similar, suggesting that the data augmentation technique was not significantly helpful to sampleRNN. The outcomes from these trainings can be divided in several classes: speech-like babbling (a), whistling (b), rumbles and impulses (c) or mixtures of the above. Figure 2 shows spectrograms of each class. Please note that the whistling tones do not appear in the output generated from the Rima Glottidis speech-only subset. It was noted, indeed, that these have the same frequency of bell tones found in the complete Rima Glottidis dataset (see e.g. Figure 3). Bell tones appear on the 5% of the Rima Glottidis dataset. This means that very simple and repetitive material can be easily reproduced by sampleRNN even though it does appear on a small portion of the dataset. The network, however, did not learn the envelope of the bell tones, as it does not replicate its immediate attack and long decay.

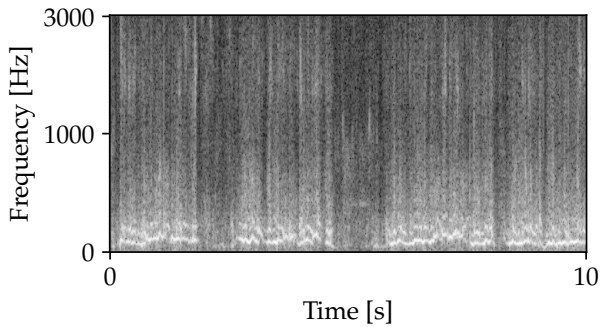
On the contrary, it is known from the experiments reported by [13] that piano tones can be generated, thus, sampleRNN is able to learn the time envelope of a class of signals, although it failed in our experiments, probably due to the lack of material.

From an artistic standpoint these samples have a relevance, especially for their unpredictability, although unfortunately they do not match the training material as one could desire.

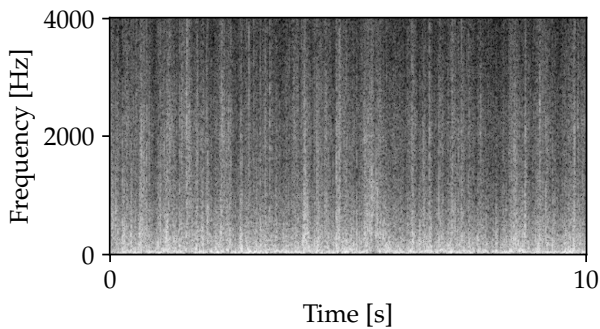
As a last note, more material was generated using the algorithm proposed in [18] for timbre transfer. For this algorithm, based on a neural autoencoder, a training set is fed to the network during training with the goal of minimizing the difference between input and output. The network retains features from the training set and should be able to transfer these on a novel input during generation. In our



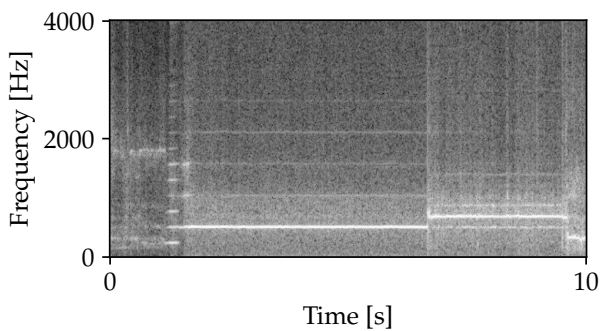
(a)



(b)

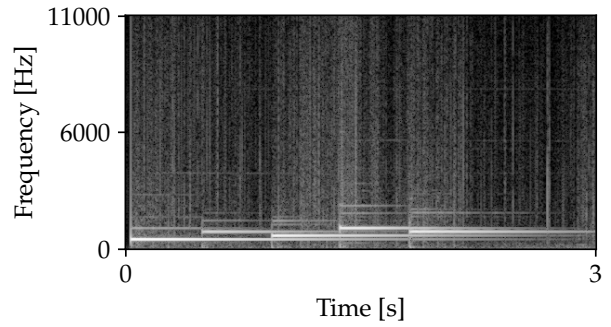


(c)



(d)

**Figure 2:** Spectrograms of several classes of outputs from sampleRNN: choir textures (a), speech-like babbling (b), rumble (c), whistle (d). The choir textures are obtained from the Chorales training set, while the others are from the Rima Glottidis training set. The main two tones in Figure (d) are a C5 and a F5, to be compared to those in Figure 3.



(a)

**Figure 3:** Spectrogram from an excerpt of bell tones from Rima Glottidis, showing tones C5 A5 F5 C6 A5.

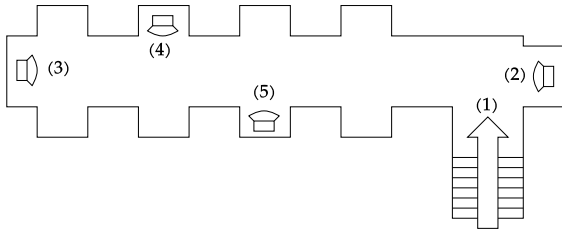
case we employed the chorales dataset for training and excerpts from Rima Glottidis during generation. The result is a vocoder-like effect that reminds of male choirs or female choirs depending whether the input voice is from a male or a female speaker, in accordance to the finding in [18] that the network maps learned frames by pitch similarity.

### 3. DESIGN OF THE INSTALLATION

Selection of the material was done on the basis of its novelty or the presence of a recognizable vocal character, since the artist's idea behind for Rima Glottidis was to investigate the voice production mechanism in unpredictable and funny ways. The audio clips were arranged to follow three phases that account for the different sorts of material produced by the network: a textural introduction where choirs are predominant and fill the ambience; an *explorative* part where speech-like sounds come from different positions of the cellar where the user is left to wander, and a last part where some freedom is taken and sounds are arranged musically. The clip is closed with a crossfade between an original excerpt of Rima Glottidis and its timbre transferred version. Some freedom was taken in using effects such as resampling, granulation and in adding small fragments of the original input material.

A site-specific spatialization was created by hiding multiple speakers in the recesses of the cellar as though creatures were hiding there with their peculiar voices. Figure 4 shows the cellar main corridor and positioning of the loudspeakers. Routing of the 4-way output audio was done on the basis of the loudspeaker localization. Two loudspeakers at the end and the beginning of the main corridor served to convey background textures and tones to fill the ambience, while the two additional loudspeakers were left for more peculiar effects, voices and noises. The user was free to walk along the corridor. The recesses hiding the loudspeakers and the audio equipment were totally dark, with dim lights along the corridor allowing visitors to walk.

The installation was premiered on 22 June 2018. On 23 June before opening of the installation a short introductory talk was held in the premises of the cellar to introduce the audience to artificial intelligence and the machine learning tools employed for the project. This was beneficial for vis-



**Figure 4:** Schematic view of the location and positioning of speakers (2-5). The entrance is denoted with (1).

itors that were able to better understand the implications of such a project.

Samples from this project are available at <https://soundcloud.com/keinreverb/sets/rima-glottidis>

#### 4. CONCLUSIONS

This paper reported the technical tools and the design choices behind the installation *Rima Glottidis*, a derivative work of the homonym musical project by artist *Økapi*. We describe current machine learning tools for generative audio and speech production and motivate the selection of sampleRNN, which was found to be the most suitable. A dataset was created from the artist stems, and data augmentation strategies were tested, however with no apparent improvement on the generation phase. The outcomes of the generation phase are described qualitatively. From our experiments the dataset size was too small to allow sampleRNN to learn to mimic the training set in a generative way. WaveGAN seems promising from this point of view, however it is designed to generate very short audio signals.

To the best of our knowledge, more rigorous methods to employ generative audio neural networks are yet to be found. As it was stated by one of the authors in a letter to the *Computer Music Journal* [20], raw audio signal modelling by deep learning may be the avenue to a new era in generative audio synthesis if the computer music community accepts this hard challenge. We hope this first work will help others in designing more advanced generative projects.

#### 5. ACKNOWLEDGMENTS

The authors would like to thank *Økapi* for providing his original material and some comments on the work. Credit goes to the Quiet Ensemble for suggesting audio spatialization and the peculiar location. Thanks to Blooming festival organizers for commissioning this project and providing support.

#### 6. REFERENCES

- [1] G. Papadopoulos and G. Wiggins, “AI methods for algorithmic composition: A survey, a critical view and future prospects,” in *AISB Symposium on Musical Creativity*, vol. 124, pp. 110–117, Edinburgh, UK, 1999.
- [2] T. J. Mitchell and D. P. Creasey, “Evolutionary sound matching: A test methodology and comparative study,” in *Machine Learning and Applications, 2007. ICMLA 2007. Sixth International Conference on*, pp. 229–234, IEEE, 2007.
- [3] P. Gibson and J. Byrne, “Neurogen, musical composition using genetic algorithms and cooperating neural networks,” in *Artificial Neural Networks, 1991., Second International Conference on*, pp. 309–313, IET, 1991.
- [4] E. R. Miranda, “Cellular automata music: An interdisciplinary project,” *Journal of New Music Research*, vol. 22, no. 1, pp. 3–21, 1993.
- [5] D. Burraston and E. Edmonds, “Cellular automata in generative electronic music and sonic art: a historical and technical review,” *Digital Creativity*, vol. 16, no. 3, pp. 165–185, 2005.
- [6] F. Colombo, S. P. Muscinelli, A. Seeholzer, J. Brea, and W. Gerstner, “Algorithmic composition of melodies with deep recurrent neural networks,” *arXiv preprint arXiv:1606.07251*, 2016.
- [7] N. Jaques, S. Gu, D. Bahdanau, J. M. H. Lombato, R. E. Turner, and D. Eck, “Tuning recurrent neural networks with reinforcement learning,” 2017. <https://openreview.net/pdf?id=Syvv2e-Kx>.
- [8] F. Fontana, “Interactive sound synthesis by the Lotka-Volterra population model,” in *Proceedings of the 19th Colloquium on Music Informatics (P. Polotti, G. Klauer, F. Fontana, and C. Drioli, eds.), (Trieste, Italy)*, pp. 24–29, 2012.
- [9] F. Fontana and V. Manca, “Discrete solutions to differential equations by metabolic P systems,” *Theoretical Computer Science*, vol. 372, no. 2-3, pp. 165–182, 2007.
- [10] A. W. Su and L. San-Fu, “Synthesis of plucked-string tones by physical modeling with recurrent neural networks,” in *Multimedia Signal Processing, 1997., IEEE First Workshop on*, pp. 71–76, IEEE, 1997.
- [11] C. Drioli and D. Rocchesso, “A generalized musical-tone generator with application to sound compression and synthesis,” in *Acoustics, Speech, and Signal Processing, 1997 IEEE International Conference on*, vol. 1, pp. 431–434, IEEE, 1997.
- [12] A. v. d. Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, and K. Kavukcuoglu, “Wavenet: A generative model for raw audio,” *white paper, available online at https://arxiv.org/pdf/1609.03499.pdf*, 2016.
- [13] S. Mehri, K. Kumar, I. Gulrajani, R. Kumar, S. Jain, J. Sotelo, A. Courville, and Y. Bengio, “SampleRNN: an unconditional end-to-end neural audio generation model,” in *5th International Conference on Learning Representations (ICLR 2017)*, 2017.

- [14] C. Donahue, J. McAuley, and M. Puckette, “Synthesizing audio with GANs,” in *International Conference on Learning Representations, ICLR 2018*, (Vancouver, Canada), 2018.
- [15] J. Engel, C. Resnick, A. Roberts, S. Dieleman, D. Eck, K. Simonyan, and M. Norouzi, “Neural audio synthesis of musical notes with WaveNet autoencoders,” *arXiv*, 2017.
- [16] Z. Jin, A. Finkelstein, M. Gautham J., and L. Jingwan, “FFTNet: A real-time speaker-dependent neural vocoder,” in *43rd IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, April 2018.
- [17] L. C. Piqueras, “Autoregressive model based on a deep convolutional neural network for audio generation.” MSc. thesis, Tampere University, Finland, 2016.
- [18] L. Gabrielli, C. E. Cella, F. Vesperini, D. Droghini, E. Principi, and S. Squartini, “Deep learning for timbre modification and transfer: An evaluation study,” in *Audio Engineering Society Convention 144*, May 2018.
- [19] J. Schlüter and T. Grill, “Exploring data augmentation for improved singing voice detection with neural networks,” in *ISMIR*, pp. 121–126, 2015.
- [20] L. Gabrielli, “Once again.” letter to the Computer Music Journal, vol 41(3), Fall 2017.

# Points of discontinuity and spatial cadences: two experimental concepts for the development of musical theory

Luca Danieli

University of Birmingham  
Dept. of Electroacoustic Music  
Birmingham (UK)  
lxd430@bham.ac.uk  
dnllcu@gmail.com

## ABSTRACT

In this article I aim to suggest a possible experimental approach to music theory for the development of new research and musical styles that connect traditional theory with the new compositional paradigms that characterize electroacoustic and computer musics. I highlight an interpretation of musical cadence as a point of discontinuity and present some preliminary results of a perceptual study conducted at the University of Birmingham on the relationship between music and space, which derives from this approach. Following the idea of cadence as a point of discontinuity, it has been possible to form a hypothesis on the possible perception of a sensation of cadence in relation to sound trajectories moving within a circular electroacoustic space. The cognitive hypothesis supposes the impossibility for the listeners to conceive a sound "object" passing "through" them, being themselves the discontinuity point of the spatial system in which they are placed. The study has involved 41 participants among experts and non-experts of electroacoustic music composition. Various sound stimuli diffused by use of multiple trajectories were presented to the participants, who were asked to evaluate their cadence level. A preliminary result reveals a tendency among electroacoustic music composers to consider sound stimuli ending at the opposite ends of a circular space as cadential.

## 1. INTRODUCTION

The contemporary research in the context of musical perception often addresses traditional theory as a solid ground for the creation of new hypotheses related to tonal music, to be confirmed or falsified within the field of cognitive research. In some cases, research has been carried out taking alternative compositional styles as a reference - e.g. atonal music -, but studies on musical expectation in these genres have rarely proved significant to day [1].

In an article titled *An Experimental Music Theory?*, Gjerdingen [2] draws attention to the need to formulate new experimental musical theories, so to provide new points of

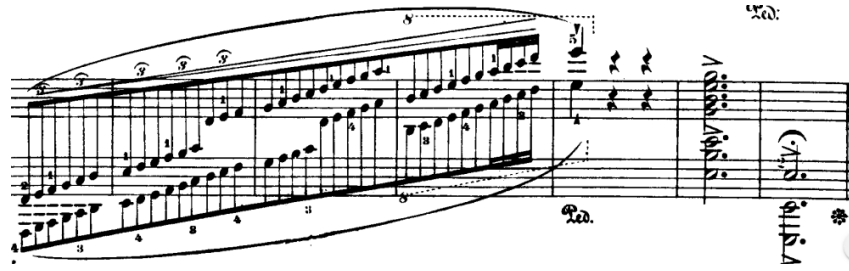
view to overcome the clichés that are still today at the basis of musical theory. Gjerdingen reports by way of example the concepts of 'tonality' and 'voice-leading', describing them as old customs often considered as incontrovertible and eternal truths that influence both the compositional activity and consequently music perception itself. Furthermore, nowadays, electroacoustic music has become an integral part of the teaching and research programs of many conservatories and universities around the world. Despite this, what is commonly conceived with the term 'musical theory' and is currently part of standard disciplinary courses on theory for music composition, is often a group of notions linked to the composition of instrumental music. Many could be the reasons of this tendency, including for me a general philosophical methodology at the ground of the electroacoustic practice and embedded profoundly in the notion of 'reduced listening', that places theorists and composers in a status of constant analysis. Thing is that after the prolific speculation on the aesthetics of electroacoustic music and its focus on descriptive methods and technological applications, contemporary music theory still struggles to deal with recent electroacoustic musical approaches.

There is therefore today the urgent need to afford alternative hypotheses for the description of common musical notions through an interdisciplinary approach that could favour the creation of new paradigms useful to open novel strategies for the study of music. The concepts of closure and cadence stand in this context as interesting objects of research, which could help advance further insight into music, musical theory and novel applications in connection with computer music and new technologies.

## 2. CADENCE AS A POINT OF DISCONTINUITY

In an article presented at the 31st session of the Musical Association, E. Markham Lee [3] has presented a concise overview on the evolution of musical closures and the development of cadence from the Middle Ages to the Romanticism, introducing how today the two notions are often considered synonyms although in origin they were indicating two independent concepts. Markham Lee has observed how the initial forms of closure used in the medieval ecclesiastical music consisted of major- and minor-second descending intervals placed at the end of a melodic fragment.

Copyright: ©2018 Luca Danieli et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution License 3.0 Unported](https://creativecommons.org/licenses/by/3.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

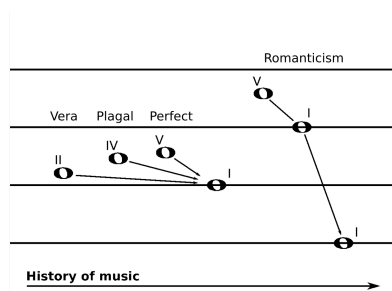


(a) A fragment from *Scherzo No. 4, Op. 54* by Chopin.

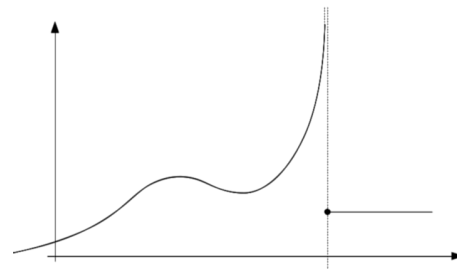


(b) A fragment from *Metastasis* by Xenakis, final closure at bar 336.

**Figure 1:** Two examples of musical closures.



(a) Evolution of closures until the Romanticism.



(b) Mathematical interpretation of a closure.

**Figure 2:** Two schematic representations of the musical closure.

Following the path traced by Markham Lee, the evolution of the musical cadence as a form of closure began with the polyphonic creation of the *Clausola Vera*, a harmonic passage consisting of a descending second in the melody accompanied by an interval ascending to the same note in the bass; passage that in those times was not yet conceived harmonically, but rather as an accidental overlapping of melodic lines. Later on, the *Clausola Plagalis* (consisting of a melodic interval of descending fourth) came to designate a compositional technique aimed at reinforcing the sense of termination of a musical passage. With the work of Monteverdi and the establishment of the harmonic reasoning during the 17th century, the Perfect Cadence (interval of descending fifth) came to be addressed as the new standard model for the termination of music, which would find further developments in the works of Bach, Handel, Haydn and Mozart, as a basic means for the creation of musical structures.

After Beethoven, it is possible to find within the Romantic music literature a new tendency to develop the musi-

cal closure by using chords (mainly Dominant and Tonic) progressively spaced out over different octaves, often repeating the same Tonic on different registers. Some examples can be found in the works of Liszt (*Hungarian Rhapsody No. 19 in D minor*), Chopin (*Piano Sonata No. 3, op. 58; Piano Sonata No. 1, op. 4; Scherzo No. 4, op. 54*), Debussy (*Pour le piano (toccata)*) and Strauss (*Also sprach Zarathustra*). An example taken from *Scherzo No. 4, op. 54* by Chopin is provided in Figure 1. In these works, it is possible to interpret the intention of the composers as a desire to break the horizontal development of voices through the introduction of one or multiple ceases, or the introduction of rests so to make the flow discontinuous. A schematic representation of this cadential development, oriented towards the progressive separation of the last chords of a composition, is shown in Figure 2a.

This tendency to fragment the melodic and harmonic flow, together with the increasing spatialization of the last chords in the music of the Classical period, can be interpreted as a point of discontinuity. This behaviour becomes more ev-

ident in the works that make great use of glissandi, as for Ferneyhough's *String Quartet No.3* or as in the case of Xenakis' *Metastasis*, in which a brief silence is inserted in the final cluster with the aim of breaking its continuity and providing the listener with a sense of conclusion. Figure 2 proposes a visual comparison between the schematic representation of the evolution of the cadence up to the Classical period and a point of discontinuity of a mathematical function. In this supposition, a musical cadence would resemble a point of discontinuity, favoring a possible reformulation of the concept within the context of computer music.

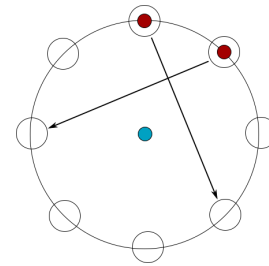
### 3. THE IDEA OF SPATIAL CADENCE

The introduction of an idea of cadence as a point of discontinuity permits to translate a concept that has originated in the context of tonal harmony into extra-musical dimensions. In particular, this interpretation has allowed to create a cognitive conjecture to study the concept of cadence in relation to a spatial context, in which tonality has proven difficult to map and is therefore impossible to use the traditional paradigms.

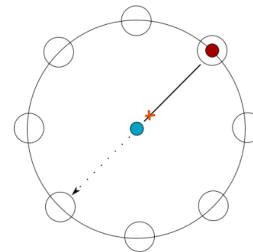
Considering an acousmatic space defined by eight loudspeakers arranged in an octagonal shape over an ideal circle whose centre is represented by the listener, the subject can be supposed to constitute the cardinal element of the system which permits the illusion of movement for the sounds projected into that space. In this perspective, one can also suppose the listener to be a blind spot of the space in which s/he is submerged, able to conceive sounds as moving around him/er but limited to conceiving spatial movements that intersect his/er position. Was there the sound of a car that gave him the illusion of moving against him, the subject would perhaps have the unconscious reaction of having to move, not being able to conceive "something" passing "through" him/er. Following this supposition, the listener would behave as the discontinuity point of the system. Assuming this as a cognitive phenomenon, and recalling the interpretation of cadence as a point of discontinuity, it is legitimate to wonder whether a sound "object" which gives the illusion of "passing through" the subject produces a sense of cadence or of closure. In particular, through the use of abstract sounds that do not represent real objects of the world, it is possible to ask whether a "cadence" would occur for each pair of loudspeakers placed at the ends of an imaginary axis passing by the centre (Figure 3).

### 4. AN EXPLORATORY STUDY ON THE PERCEPTION OF SPATIAL CADENCES

An exploratory study on the possible existence of a sensation of "cadence" for sonic stimuli moving in space has been conducted at the Electroacoustic Music Studios of the University of Birmingham between May 2016 and December 2016. The studies that have been carried out to date on the perception of cadences and closures include research related to harmony [4, 5], melody [6], morphology [7], tonality [8], tones and timbres [5]. Studying the



(a) Trajectories perceived as possible.



(b) Impossible trajectory: point of discontinuity.

Figure 3: Cognitive hypothesis of a spatial cadence.

ability to perceive the musical closures of different musical styles and supposing the segmentation ability as a prerequisite to detect closures within a piece, Peebles [7] has presented to listeners some fragments taken from the string quartet repertoires of Mozart and Bartók, asking the participants to mark the points of segmentation of the fragments played. His experiment has shown that the segmentation ability is influenced by both the familiarity of the listeners with the musical genres heard and the musical preparation of the participants. More experienced musicians are able to adapt more quickly to the musical genre in analysis, so as to facilitate their faculty of segmentation and the consequent perception of musical closures. Sears, Caplin and McAdams [4], in an attempt to study the cognitive mechanisms underlying the perception of cadences in tonal music, have presented to their participants 50 fragments taken from Mozart's piano sonatas, asking them to rate the level of completion for each fragment. The study has shown a difference between musicians and amateurs in rating different types of cadence, making musicians more sensitive to the harmonic variations listened. To day, the concept of cadence has not been studied in relation to spatial movements. Music has found application in the spatial field specifically in connection with the auditory perception, with the aim of studying the localization of sounds around the listener [9]. For listeners trained in the localization of sound events, it is believed that the accuracy of localization varies from 2-3 degrees, for sound stimuli coming from the front, to about 15 degrees for sounds coming from the back area [10]. In three experiments, Recanzone, Makhmra and Guard [11] study the accuracy of localization for sounds with different spectra, detecting an average localization of 4 degrees for white noise and 8-10 degrees for stimuli with a frequency of 1 kHz or higher. Despite the reduced number of studies connecting music and space, Sievers et al. [12] have highlighted a correlation

between music and spatial movements in an experiment aimed at studying the effect of audiovisual parameters on emotions. By splitting participants into two groups that could not interact, the first group was presented synthetic piano melodies that were to be modified through the manipulation of 5 faders; the second group was instead presented the animation of a bouncing ball, whose behaviour was modifiable through the same faders. These 5 faders indicated common parameters, named "rate", "jitter", "consonance/smoothness", "step-size" and "direction". The algorithm leading the behaviour of both stimuli was using the same probabilistic model in both experiments. All participants were asked to manipulate the available parameters in relation to five suggested emotions, namely "angry", "happy", "peaceful", "sad", and "scared". The experiment has shown a similarity in the manipulation of the parameters in both groups, leading researchers to hypothesize that space and music might have a common universal structure. The experiment following described is exploratory and wants to emphasize the assumption advanced by Sievers et al. [12], trying to understand whether it is possible to trace a sensation of "cadence" for sonic trajectories based on the conjecture presented above.

## 4.1 Methodology

### 4.1.1 Participants

The study has counted 41 participants, aged 19 to 60, divided into 3 groups.

- Amateurs: 18 participants who do not play instruments and do not know the concept of cadence;
- Performers of traditional instruments (performers): 12 participants with experience in the use of musical instruments and knowledge of the concept of cadence, but without experience in manipulating spatial parameters;
- Composers of electroacoustic music (composers): 11 participants, familiar both with the concept of cadence and with the manipulation of spatial parameters. This group includes professional composers and students with at least 1 year of experience in manipulating spatial sounds.

### 4.1.2 Stimuli

4 types of stimuli have been used in the experiment. Three stimuli (**a**, **b**, and **c**) are composed of 5 impulses, while one stimulus (**d**) is made of 17 equidistant impulses. Two stimuli (**a** and **d**) present pulses with a duration of 100ms each and white spectra; the stimulus (**b**) is obtained from the synthesis of 9 harmonic sinusoids on a fundamental of 2001 Hz, and the impulses follow one another repeatedly at distances of 600ms and 300ms; the stimulus (**c**) presents 5 piano notes at frequencies of [1920, 2155, 2419, 1812, 1920] Hz (equivalent to a transposition of [Do, Re, Mi, Si, Do]), and the impulses follow one another repeatedly at distances of 550ms and 275ms. The duration of the stimuli varies between 1050 and 2400 ms overall.

### 4.1.3 Environment

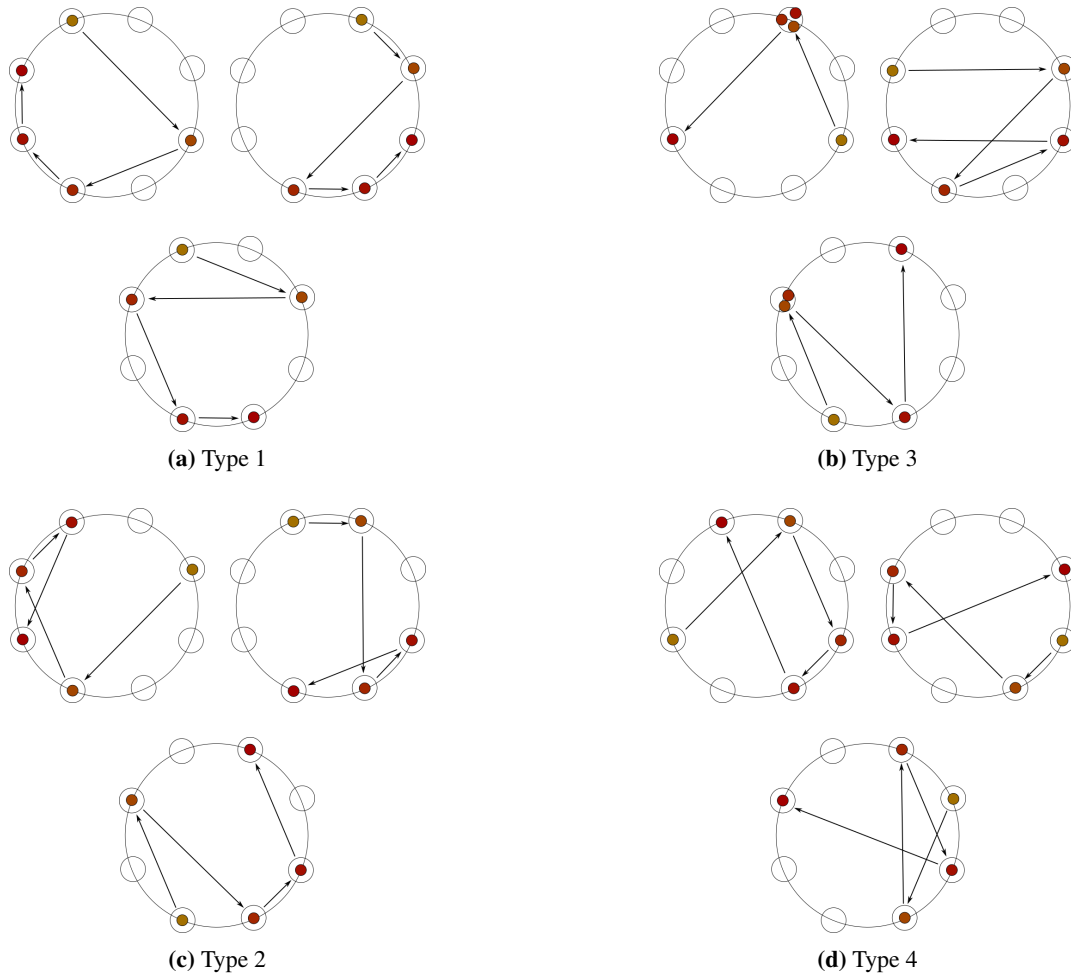
The experiment was conducted in a recording room of the Electroacoustic Music Studios of the University of Birmingham. The stimuli were reproduced through the use of 8 Genelec 8030A, placed at a distance of about 180cm from the listener. The loudspeakers have been rotated of 22.5 degrees from the vertical axis, so as not to coincide with the listener's cardinal points.

### 4.1.4 Process

Every experiment is composed of 8 consecutive tests. Each test presents 1 type of stimulus, spatialized by use of 25 different trajectories. Some examples of these 25 trajectories are shown in Figure 4. The order of the trajectories is randomized in each test; furthermore, all the trajectories are randomly rotated and mirrored so as to avoid correlations between the listener's orientation and the orientation of the trajectories presented. The 8 tests are grouped into 2 groups; both groups present 4 tests, each with a different type of stimulus (**a**, **b**, **c**, and **d**). The two groups follow one another without interruption, in order to give the listener time to familiarize with the environment. The trajectories are spatialized through the use of 5 loudspeakers, with no multiple impulses being played simultaneously. (**a**, **b**, and **c**) present 1 pulse per loudspeaker, while (**d**) presents 4 pulses per loudspeaker except for the last one. The 25 trajectories in each test have 4 types of closures. Closures of the first type locate the last two impulses of the trajectory on adjacent loudspeakers; those of the second type locate the last two impulses on loudspeakers placed orthogonally with respect to the listener; those of the third type locate the last two impulses at an angle of 135 degree; those of the fourth type locate the last two impulses at 180 degree, in opposite directions from the listener. Each participant was asked to rate the cadence level of each trajectory on a 5-point Likert scale, in a time interval of 5 seconds between one stimulus and the next. Since amateurs did not know the concept of cadence, they were provided with some synonyms such as "closure", "completion" and "satisfaction" as potential parameters for the rating of the trajectories before the beginning of the experiment.

## 4.2 Analysis

For the analysis of the experiment, it has been chosen to use a linear mixed-effect model using the R software. The analysis model was identified by comparison, introducing new conditions progressively and comparing the models through a Chi-Squared test. The Chi-Squared test is reported in Table 1. The random effects that have been considered in the analysis model include the by-subject random intercepts and the by-item random intercepts. The by-subject random intercepts represent the average difference in assessment among the participating subjects. The by-item random intercepts represent the average rating for each type of closure (type 1, 2, 3, 4) for each subject. Since each closure typology is reproduced through multiple trajectories that have different properties, each participant could have a preference for a specific trajectory in that category (due to factors that are not related directly to

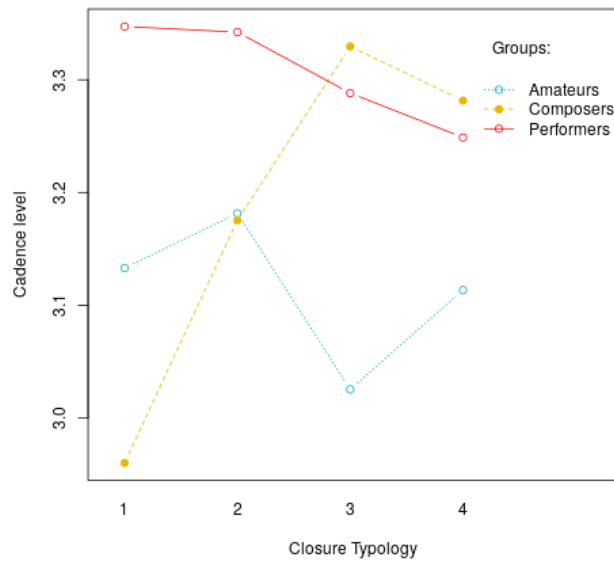


**Figure 4:** 12 trajectories used to spatialize the sonic stimuli during the experiment.

|          |   |
|----------|---|
| Models:  |   |
| Model 1: | Score ~1 + (1 Participant) + (1 Item)                       |
| model 2: | Score ~Closure + (1 Participant) + (1 Item)                 |
| model 3: | Score ~Closure + ExpertiseType + (1 Participant) + (1 Item) |
| model 4: | Score ~Closure * ExpertiseType + (1 Participant) + (1 Item) |

| Model | DF | AIC   | BIC   | logLik | deviance | Chisq Chi | Df | Pr(>Chisq)    |
|-------|----|-------|-------|--------|----------|-----------|----|---------------|
| 1     | 4  | 24694 | 24722 | -12343 | 24686    |           |    |               |
| 2     | 6  | 24697 | 24739 | -12343 | 24685    | 1.2965    | 2  | 0.52297       |
| 3     | 9  | 24697 | 24759 | -12339 | 24678    | 6.9578    | 3  | 0.07326 .     |
| 4     | 15 | 24669 | 24774 | -12320 | 24639    | 39.1579   | 6  | 6.665e-07 *** |

**Table 1:** Chi Squared test for choosing the analysis model.



**Figure 5:** Interaction between the groups of participants and the typologies of closure.

the type of closure itself), thus offsetting the ratings within each category and consequently the preferences for each participant. The analysis of the model has been carried out through the creation of contrast masks for planned comparison, in which it was assumed that closures of the fourth type would have a prevailing effect. The two main effects considered in this model include the type of closure and the group of participants.

### 4.3 Results

The results have highlighted an interaction between the groups participating and the types of closure ( $p < .001$ ), and a main effect of the types of closure within the group of composers ( $p < .001$ ), suggesting that the perception of cadence is for the composers of electroacoustic music in correlation with the type of closure implemented in the trajectory listened to. Closures of the fourth type have been found to be statistically more effective than those of the second ( $p < .01$ ) and first type ( $p < .001$ ). Furthermore, closures of the third and fourth type were observed to be completely more effective than those of the second and first type ( $p < .001$ ), suggesting the tendency to consider distant closures as more cadential. Therefore, it has not been possible to detect a statistical difference between closures of the third and fourth type, and the tendency to consider both of them as more cadential may be related to the inaccuracy of sound localization described before. Closures of the first type, instead, were found to be statistically less effective than those of the second ( $p < .05$ ), third ( $p < .001$ ) and fourth type ( $p < .001$ ).

## 5. CONCLUSIONS

Overall, this study highlights two very promising results for electroacoustic and computer music. The first result

consists in an alternative interpretation of the concept of musical cadence as a point of discontinuity. One of the major problems in the research field of contemporary music theory is the need to hypothesize new scenarios that can overcome the traditional models of tonal theory, and the conception of a musical cadence as a point of discontinuity could lead towards the opening of new fields of research especially in connection with computer music and mathematics. The second result consists in the detection of a perceptive phenomenon that can help to develop the theory of music in connection with space and the generation of new approaches for the composition of electroacoustic music, use musical concepts as a means to discover new spatial properties, and find further applications in connection with dance and those musical technologies that use space as a controller. This experiment has indeed highlighted an interaction between experts of electroacoustic music trained in spatial music listening and the other types of participants, which might suggest that future research could provide new insight into this phenomenon.

### 5.1 Future research

The results that have been obtained from this preliminary analysis of the experiment seem to point towards a confirmation of the conjecture advanced from a musical point of view, but have been unable to provide a clear feedback on the cognitive hypothesis for which it would be impossible for the listener to conceive a sonic object moving "through" him/er. This may be due to either the use of point-source impulses, the lack of transitions between different loudspeakers or the abstract nature of the presented stimuli. Future research might point towards different directions, providing either new insight into the cognitive hypothesis (by use of realistic sounds that could reinforce the appearance of the supposed phenomenon or change the listening

mode of the subjects involved), or new research on the musical perception of space. The study herein presented has been exploratory and has therefore included a large number of variables in order to study the matter from multiple perspectives, which are now under analysis. Given thus its original goal, the results have been beyond expectation, and they permit to hypothesize that further studies might improve this research by means of more specific designs.

## 6. ACKNOWLEDGMENTS

I would like to thank Dr. Maria Witek for advising me on the best strategies for the analysis of the data.

## 7. REFERENCES

- [1] A. Ockelford and D. Sergeant, “Musical expectancy in atonal contexts: Musicians’ perception of “antistructure”,” *Psychology of Music*, vol. 41, no. 2, pp. 139–174, 2012.
- [2] R. O. Gjerdingen, “An experimental music theory?,” in *Rethinking music* (N. Cook and M. Everist, eds.), pp. 161–170, Oxford University Press, 1999.
- [3] E. Markham Lee, “Cadences and closes,” in *Proceedings of the Musical Association, 31st Sess.*, pp. 63–68, 1904.
- [4] D. Sears, W. E. Sergeant, and S. McAdams, “Perceiving the classical cadence,” *Music Perception*, vol. 31, no. 5, pp. 397–417, 2014.
- [5] A. J. Milne, “A psychoacoustic model of harmonic cadences,” in *7th Triennial Conference of European Society for the Cognitive Sciences of Music, ESCOM 2009*, (Jyväskylä, Finland), pp. 328–337, 2009.
- [6] P. G. Vos and D. Pasveer, “Goodness ratings of melodic openings and closures,” *Perception and Psychophysics*, vol. 64, pp. 631–639, 2002.
- [7] C. Peebles, *Role of Segmentation and Expectation in the Perception of Closure*. PhD thesis, Florida State University: College of Music, Tallahassee, 2011.
- [8] N. Cook, “The perception of large-scale tonal closure,” *Music Perception*, vol. 5, no. 2, pp. 197–205, 1987.
- [9] B. Tversky, “Spatial cognition: Embodied and situated,” in *The Cambridge handbook of situated cognition* (P. Robbins and M. Aydede, eds.), pp. 201–216, Cambridge, UK: Cambridge University Press, 2009.
- [10] S. Carlile, P. Leong, and S. Hyams, “The nature and distribution of errors in sound localization by human listeners,” *Hearing Research*, vol. 114, pp. 179–196, 1997.
- [11] G. Recanzone, S. D. D. R. Makhmra, and D. C. Guard, “Comparison of relative and absolute sound localization ability in humans,” *The journal of the Acoustical Society of America*, vol. 103, pp. 1085–1097, 1998.
- [12] M. C. B. Sievers, L. Polanky and T. Wheatley, “Music and movement share a dynamic structure that supports universal expression of emotion,” in *Proceedings of the National Academy of Sciences*, vol. 110, pp. 70–75, 2013.

# Vexations: La Serie. Dati dal corpo, interpretazione musicale e spazio d'ascolto.

Adriano Castaldini

Conservatorio di Musica “Giuseppe Tartini” di Trieste, Dipartimento di Nuovi Linguaggi Percussioni e Tecnologie Musicali, Scuola MNT (Musica e Nuove Tecnologie)  
adrjork@gmail.com

## ABSTRACT

*Vexations: La Serie* è il titolo di una performance per pianoforte e sistema sensoristico, ideata e realizzata da Adriano Castaldini nel settembre 2016. La performance è basata sull'esecuzione integrale di *Vexations* di Erik Satie – brano pianistico della durata record di circa 24 ore – in un'interpretazione aumentata dall'utilizzo di sensori EEG, EMG, GSR e di temperatura indossati dal pianista durante l'esecuzione. *Vexations: La Serie* si propone come esperimento d'interpretazione musicale in cui la particolare condizione psicofisica del pianista, in un'esecuzione di così lunga durata, influenza il timbro sonoro del pianoforte (microfonato e amplificato) grazie alla gestione informatica di una piattaforma interattiva programmata ad hoc, in grado di convertire in tempo reale i dati psicofisici rilevati, in valori di controllo dello spettro sonoro del pianoforte, pertanto rendendo udibili le vessazioni del pianista. Questa ricerca ha indagato l'efficacia artistica dei dati provenienti dal corpo fisico dell'interprete come aumentazione non già dello strumento musicale, quanto dei mezzi interpretativi, e la conseguente influenza di questi sulla forma d'ascolto (ovvero lo spazio fisico d'ascolto e le modalità di fruizione da parte del pubblico).

## 1. VEXATIONS DI ERIK SATIE

### 1.1 La composizione originale

*Vexations* (Figura 1) è un brano scritto presumibilmente per pianoforte da Erik Satie, e costituito da un *Thème* monodico di 19 note più una pausa della durata di 13 quarti, scritto su un singolo pentagramma in chiave di basso e senza stanghette di battuta, a cui seguono – secondo le indicazioni annotate dall'autore sul mano-

scritto originale<sup>1</sup> – una prima armonizzazione del tema su un sistema di due pentagrammi, una ripetizione del tema, e infine una seconda armonizzazione ancora su due pentagrammi. Con una nota<sup>2</sup>, l'autore avverte che l'intero *motif* (tema, prima armonizzazione, tema, seconda armonizzazione) va ripetuto 840 volte. L'autore non fornisce un'indicazione metronomica precisa, limitandosi ad annotare *Très lent* in partitura, che tuttavia trova corrispondenza nel range metronomico 40–44 battiti per minuto. A questa velocità, l'intero *motif* si esaurisce in quasi un minuto e mezzo (considerando i respiri e le fluttuazioni agogiche dell'esecutore), e l'intera composizione (le 840 ripetizioni del *motif*) in quasi 24 ore, durata che, confrontando le varie esecuzioni avvenute in passato, oscilla considerevolmente a seconda delle scelte di tempo e di agogica dei diversi interpreti.

Vexations

*Pour se jouer 840 fois de suite ce motif, il sera bon de se préparer au préalable, et dans le plus grand silence, par des immobilités sérieuses.*

The figure displays a semi-diplomatic transcription of Erik Satie's 'Vexations'. It consists of three systems of musical notation. The first system shows the original 'Thème' on a single bass clef staff, marked 'Très lent'. The second system shows the first harmonization on two staves (treble and bass clefs). The third system shows the second harmonization, also on two staves. Above the first system, there is a French instruction: 'Pour se jouer 840 fois de suite ce motif, il sera bon de se préparer au préalable, et dans le plus grand silence, par des immobilités sérieuses.' Below the second system, another instruction reads: 'A ce signe il sera d'usage de présenter le thème de la Basse:'. Below the third system, the word 'Thème' is written above the bass clef staff.

Figura 1. *Vexations: trascrizione semi-diplomatica del manoscritto* (© 2018 A. Castaldini).

Copyright: © 2018 Adriano Castaldini. This is an open-access article distributed under the terms of the [Creative Commons Attribution License 3.0 Unported](https://creativecommons.org/licenses/by/3.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

<sup>1</sup> Il manoscritto, appartenuto inizialmente alla collezione Claude Rostand, è successivamente passato agli archivi della Fondation Erik Satie a Parigi, ed infine all'IMEC (Institut Mémoires de l'édition contemporaine) nel 2000.

<sup>2</sup> «Pour se jouer 840 fois de suite ce motif, il sera bon de se préparer au préalable, et dans le plus grand silence, par des immobilités sérieuses.»

## 1.2 La prima esecuzione in staffetta

*Vexations* è stato composto nell'ultima decade del XIX secolo, ma stampato soltanto nel 1949 (facsimile del manoscritto originale [1]) grazie al compositore americano John Cage, quindi pubblicato nel 1958 (ancora facsimile [2]), ed eseguito per la prima volta nel 1963<sup>3</sup> da una staffetta pianistica (J. Cage incluso), ogni pianista eseguendo solo un ristretto numero di ripetizioni. Dal '63 ad oggi, *Vexations* è stato eseguito molte volte, il più delle quali in staffetta, e molto raramente in versione solistica.

In realtà, sulla partitura manoscritta Satie annota: «Per suonare il tema 840 volte di seguito, sarebbe opportuno prepararsi in anticipo, e nel più profondo silenzio, con seriosa compostezza», pertanto tutte le esecuzioni staffettistiche, da quella storica di J. Cage in poi, contravvengono chiaramente l'esplicita indicazione dell'autore, mutando (e certamente impoverendo<sup>4</sup>) la concezione originale del brano. È da notare, però, che rispetto all'arbitrio staffettistico, l'esecuzione di Cage non apporta alcuna innovazione relativa allo spazio fisico d'ascolto<sup>5</sup>.

## 1.3 Le esecuzioni solistiche

Secondo le informazioni raccolte da Gavin Bryars<sup>6</sup>, la prima esecuzione di *Vexations* completa e “in solitaria” – quindi filologicamente corretta – è di Richard Toop, ed è avvenuta a Londra nel 1967. Da allora, le esecuzioni solistiche saranno piuttosto rare e corrispondono alle première nazionali – questo sia per la difficoltà di legare un'esecuzione di *Vexations* dalla logica dell'evento (che tende a perdere “senso” nel replicarsi senza significative varianti), che per lo sforzo psicofisico che un'integrale solistica richiede. Infatti, in alcune esecuzioni integrali più recenti<sup>7</sup>, lo stato psicofisico del pianista è stato monitorato in tempo reale con equipaggiamento medico<sup>8</sup> o, indirettamente, tramite dispositivi MIDI<sup>9</sup>. Tuttavia è fondamentale, ai

fini della presente ricerca, precisare che in queste esecuzioni il monitoraggio psicofisico è sempre rimasto estraneo alla performance artistica in sé, utilizzato esclusivamente per analisi scientifica.

## 2. VEXATIONS: LA SERIE

### 2.1 Ipotesi di ricerca

L'ipotesi alla base di questa ricerca è che l'annotazione di Satie non soltanto contraddica le esecuzioni staffettistiche e l'idea interpretativa di J. Cage, ma allontani *Vexations* dalla concezione (e fruizione) tardo ottocentesca di brano da concerto, avvicinandolo altresì ad un progetto di body art ante litteram: si tratta infatti di una performance solistica concepita per vessare il performer stesso, forzandolo all'immobilità e alla veglia, nonostante un'ipnosi musicale lunga un giorno intero. Dunque una sfida che richiede all'interprete un adeguato allenamento psicofisico, prima ancora che musicale. La performance *Vexations: La Serie* è quindi pensata come ricerca pratica tesa a vagliare l'efficacia artistica di questa ipotesi, valutando la necessità di un'aumentazione tecnologica dei mezzi interpretativi (non semplicemente strumentali) e di un conseguente ripensamento dei tradizionali spazi d'ascolto<sup>10</sup> per il pubblico.

### 2.2 Metodo

La performance *Vexations: La Serie* è stata ideata, realizzata ed eseguita da Adriano Castaldini, studente di Composizione Videomusicale<sup>11</sup> presso il Conservatorio “Giuseppe Tartini” di Trieste, con la supervisione del prof. Pietro Polotti<sup>12</sup>, ed ha avuto luogo nel settembre 2016 in Piazza Unità d'Italia, in occasione delle tre giornate del festival scientifico–tecnologico Trieste Next.

A differenza delle precedenti esecuzioni “monitorate”, questa è stata, in assoluto, la prima ad aver uti-

<sup>3</sup> *Pocket Theatre*, New York.

<sup>4</sup> Tuttavia, nel giudizio complessivo rispetto all'arbitraria première cageiana, è inevitabile tener conto della particolare ricerca artistica del compositore americano, che comprende Minimal Music e filosofie orientali – e in quest'ottica *Vexations* può effettivamente essere immaginato come esempio prodromico di un possibile “Zen for Music” (parafrasando Nam June Paik: l'opera *Zen for Film* è stata infatti realizzata dal video-artista sudcoreano un anno prima della staffetta dell'amico J. Cage).

<sup>5</sup> Come forse sarebbe stato prevedibile, pensando che nel 1973 Cage avrebbe eseguito la sua epitome, *4'33"* (1952), all'aperto in Harvard Square.

<sup>6</sup> L'articolo di Gavin Bryars del 1979 *Vexations and its performers* [3] contenente l'elenco di tutte le esecuzioni pubbliche note di *Vexations* fino a quel momento, avrebbe dovuto far parte di una monografia dedicata a Satie per la serie *Les Cahiers de l'Herne*, che però non fu mai pubblicata. Il testo in inglese è stato poi pubblicato nel 1983 in *Contact* n. 26.

<sup>7</sup> S'intendono qui le esecuzioni successive al 1979, anno di stesura dell'articolo di Gavin Bryars.

<sup>8</sup> Pianista Silvio Feliciani. L'Aquila, 1988.

<sup>9</sup> Secondo Reinhard Kopiez [4], i dati agogici e dinamici raccolti durante la performance di Armin Fuchs su Disklavier (Dresda, 2000)

permettono di riconoscere la fase di trance occorsa al pianista durante l'esecuzione.

<sup>10</sup> In realtà lo spazio d'ascolto è stato rivoluzionato più volte nella storia della musica – dai teatri greci all'aperto, alle chiese, ai salotti, alle sale da concerto, ai dispositivi domestici e portatili di riproduzione audio – ed è stato tanto influente sulla forma di alcune composizioni (basti pensare ai brani con cori “battenti” dei Gabrieli, che nel XVI secolo sfruttavano acusticamente le caratteristiche cantorie contrapposte nella basilica di San Marco a Venezia) quanto altresì influenzato da alcune “concezioni compositive” (l'innovativo Festspielhaus di Bayreuth voluto da Wagner per il suo *Der Ring des Nibelungen*, o l'utopistico teatro himalayano per *Mysterium* di Skrjabin, fino ad arrivare – in “epoca elettronica” – all’“arca” di Renzo Piano per *Prometeo* di Luigi Nono). Tuttavia, da un punto di vista squisitamente interpretativo – soprattutto per il caso qui trattato di un brano tardo-ottocentesco – l'influenza esercitata dall'interpretazione sullo spazio fisico d'ascolto è solitamente trascurabile (semmai è l'interprete ad “adattarsi”, subendo/sfruttando le caratteristiche fisiche dello spazio che ospita la sua esecuzione). Per questo, la ricerca di un'influenza sullo spazio fisico d'ascolto in *Vexations: La Serie* rappresenta una novità nel panorama interpretativo del repertorio pianistico.

<sup>11</sup> Biennio Specialistico di Musica e Nuove Tecnologie.

<sup>12</sup> Dipartimento di Nuovi linguaggi, percussioni e tecnologie musicali.

lizzato il monitoraggio psicofisico dell'esecutore come evento sonoro–musicale e come chiave interpretativa del brano.

Il metodo utilizzato consiste in un sistema sensoristico indossato dal performer (sensori EEG<sup>13</sup>, EMG<sup>14</sup>, GSR<sup>15</sup> e di temperatura) messo in relazione ad un sistema microfonic e di amplificazione applicato al pianoforte. La relazione tra i due “corpi” (del performer e del pianoforte) è ottenuta grazie ad una piattaforma interattiva in grado di trasformare i dati rilevati dai sensori in tempo reale in valori di elaborazione del suono del pianoforte amplificato. Il decorso delle reazioni psicofisiche del performer durante l'esecuzione determina quindi i mutamenti continui delle caratteristiche spettrali del suono riprodotto dagli altoparlanti, e pertanto rientra completamente nell'azione artistica. Questo avvalorava l'ipotesi di poter concepire e proporre *Vexations* come progetto di body art ante litteram: se il senso della body art è (generalizzando) rendere “visibili” le reazioni del corpo, in *Vexations: La Serie* reazioni fisiche e mentali sono rese udibili, pertanto parte integrante la performance in sé.

Durante tutta la performance, l'esecutore ha un feedback visivo dei suoi dati psicofisici, ed è conscio delle relazioni programmate tra sensori ed elaborazione del suono (in questo caso, l'esecutore è anche il programmatore).

## 2.3 Mezzi

I mezzi utilizzati in questo progetto sono principalmente: pianoforte, sensori, microfoni, scheda audio, altoparlanti, computer, schermo secondario, piattaforma interattiva e superfici di controllo MIDI.

### 2.3.1 Pianoforte

Per la performance pubblica è stato utilizzato un pianoforte Yamaha C5 noleggiato dal Conservatorio, ma per i primi test domestici sono stati utilizzati sia un C5 che uno Steinway B211. Lo scopo di utilizzare pianoforti diversi era capire se strumenti di grandezza simile (entrambi  $\frac{3}{4}$  di coda), ma non dello stesso marchio, potessero influenzare diversamente il sistema microfonic–sensoristico, e di conseguenza capire come eventualmente uniformare la risposta del sistema attraverso regolazioni nella piattaforma interattiva. Si è quindi dotata la piattaforma di sistemi di autoregolazione del volume, utili anche a prevenire l'effetto Larsen.

Sempre per prevenire l'effetto Larsen in esecuzione, e per attutire il suono dello strumento a favore del suono elaborato proveniente dagli altoparlanti, si è scelto di tener chiuso il coperchio del pianoforte<sup>16</sup> e di coprire lo strumento con del panno.

<sup>13</sup> Elettroencefalografici.

<sup>14</sup> Elettromiografici.

<sup>15</sup> Elettrodermici.

<sup>16</sup> È stato necessario utilizzare microfoni con forme specifiche per poter essere piazzati nel pianoforte chiuso (sez. 2.3.3). Sono stati

La tipologia del pianoforte – a coda anziché verticale – è stata scelta per agevolare il piazzamento dei microfoni sul telaio di ghisa.

È infine importante notare che la performance non prevede l'uso dei pedali del pianoforte per non confondere le modificazioni elettroniche del suono con gli effetti strumentali di risonanza e di *una corda*. Pertanto, allo scopo di mantenere il legato pur senza il pedale di risonanza nelle parti accordali del *motif* (le due parti armonizzate), è stata studiata una diteggiatura speciale che sfrutta il solo legato di dito (anche prevedendo lo scambio di note tra mano destra e sinistra). Inoltre, la diteggiatura prevede l'esecuzione del *Thème* monodico all'inizio del *motif* con la mano sinistra, e la sua ripetizione (dopo la prima armonizzazione, a metà del *motif*) con la mano destra: questo espediente permette di capire rapidamente in quale punto del *motif* ci si trovi, anche quando, durante la performance, si sia entrati in uno stato di dormiveglia o addirittura di trance<sup>17</sup>.

### 2.3.2 Sensori

Lo scopo di utilizzare diversi tipi di sensori (EEG, EMG, GSR e di temperatura) è catturare lo stato psicofisico del performer nella sua complessità e varietà di risposte.

Come sensoristica EEG è stato scelto il casco EPOC+ di Emotiv, dotato di 14 elettrodi collocati simmetricamente in AF3/AF4, F3/F4, F7/F8, FC5/FC6, T7/T8, P7/P8, O1/O2 (secondo il Sistema 10–20<sup>18</sup>, Figura 2). Si tratta di un sistema di sensori umidi a soluzione salina, adatti a sessioni d'utilizzo prolungate<sup>19</sup>. Il casco comunica con il computer via wi-fi tramite un ricevitore USB. Grazie al Community SDK<sup>20</sup> e all'aiuto del prof. Stefano Bonetti<sup>21</sup> è stata

---

messi alcuni spessori in gomma tra il coperchio e la cassa del pianoforte per non pizzicare i cavi microfonic.

<sup>17</sup> Secondo Reinhard Kopiez [4], se si seguono le indicazioni di Satie, è impossibile eseguire *Vexations* senza entrare in trance. In realtà quest'evenienza non si è verificata nella performance triestina del 2016, e il motivo è probabilmente la diversa forma d'ascolto prevista per il pubblico: la possibilità per gli ascoltatori di muoversi e di parlare al pianista durante l'esecuzione (sezioni 2.5, 2.7 e 2.8) ha costituito per quest'ultimo uno stimolo attentivo continuo.

<sup>18</sup> Il *Sistema internazionale 10–20* è un posizionamento standard dei sensori EEG sullo scalpo. I sensori sono collocati a distanze del 10% o del 20% rispetto al 100% che rappresenta la distanza tra *inion* (prominenza alla base dell'osso occipitale) e *nasion* (attaccatura superiore del naso). Le sigle indicano la regione di corteccia monitorata (frontale, centrale, parietale, temporale, occipitale) e l'emisfero (numeri dispari per l'emisfero sinistro, e pari per il destro).

<sup>19</sup> In realtà, i sensori secchi sono molto più pratici, in quanto non richiedono alcuna preparazione, ma in lunghe sessioni d'utilizzo risultano essere molto irritanti per il cuoio capelluto. I sensori a gel sono altresì inadatti alle lunghe sessioni, in quanto non è possibile sostituire il gel, una volta seccato, senza staccare l'elettrodo dalla pelle introducendo una temporanea interferenza. Di contro, i sensori a soluzione salina possono essere continuamente irrorati. L'irrorazione garantisce il funzionamento virtualmente illimitato dei sensori e offre un certo sollievo al cuoio capelluto, ma implica la presenza di un assistente che durante la performance umetti i sensori con un dosatore ogni ora circa.

<sup>20</sup> Community SDK è un kit di sviluppo software disponibile gratuitamente dal sito <https://www.emotiv.com>

<sup>21</sup> Conservatorio “Giuseppe Tartini” di Trieste.

compilata un'applicazione in linguaggio C++ che invia i dati dal casco a qualsiasi altro software condivida il protocollo OSC<sup>22</sup>. I dati si aggiornano ogni mezzo secondo, e comprendono il livello di carica della batteria, la qualità di connessione di ogni singolo elettrodo, e i valori di potenza di cinque bande di frequenza<sup>23</sup> per ogni elettrodo: Theta, Alpha, Low-Beta, High-Beta, Gamma (ogni elettrodo rileva sempre tutte le bande contemporaneamente, ma è possibile valutare la loro potenza separatamente e stabilire quale sia la predominante in un dato momento per ciascun elettrodo).

Come sensoristica EMG, GSR e di temperatura è stata scelta l'unità e-Health Sensor Shield V2.0<sup>24</sup> di Cooking Hacks<sup>25</sup> montata su piattaforma Arduino<sup>26</sup>. Durante la performance del 2016, i sensori GSR e di temperatura erano posizionati sotto le dita del piede sinistro, e il sensore EMG sul muscolo bicipite brachiale, mentre l'unità e-Health/Arduino era fissata alla cintura e collegata al computer con un cavo USB da 3 metri per lasciare margine di movimento al performer.

### 2.3.3 Microfoni, scheda audio e altoparlanti

Nella performance in oggetto, i microfoni sul pianoforte hanno una funzione concettualmente simile ai sensori sul corpo: lo scopo di utilizzare diversi tipi di microfono per le diverse zone dello strumento è catturarne la varietà timbrica. Sono stati utilizzati tre tipi di microfono: 2 side-address, 2 a clip, e 3 a contatto, per un totale di 7 microfoni. La loro forma consente di chiudere il coperchio del pianoforte limitando il suono proveniente direttamente dallo strumento, in favore del suono elaborato proveniente dagli altoparlanti.

Il posizionamento dei microfoni è stato scelto empiricamente per enfatizzare la varietà timbrica dello strumento: il 1° side-address, montato con un morsetto sul telaio in ghisa all'estremo della coda e alla sinistra della corda più bassa verso la cassa armonica, ottiene un suono un po' nasale, metallico e chiuso; il 2° side-address, montato sul telaio e rivolto ai martelletti della regione centrale, evidenzia i rumori della meccanica; il 1° microfono a clip, montato sul telaio con un morsetto e orientato verso la cassa armonica dal primo foro della ghisa, ottiene un ottimo suono sebbene leggermente "schiacciato"; il 2° microfono a clip, montato sulla regione centrale del telaio e schiacciato contro il coperchio chiuso (come un microfono a contatto) rivela un suono molto aperto e bilanciato; il 1° microfono a contatto, attaccato alla serratura del pianoforte accanto al catenaccio, suona similmente ad uno strumento a corda pizzicata; il 2°

microfono a contatto, sul piccolo piatto di legno sotto la coda, cattura "l'effetto vinile"; il 3° microfono a contatto, sotto l'asta del pedale di risonanza, cattura una sorta di "effetto clavicembalo"<sup>27</sup>.

La scheda audio utilizzata è stata una MOTU 896HD firewire alla quale sono stati collegati i 7 microfoni tramite attacchi XLR.

Gli altoparlanti utilizzati nei test domestici sono stati due Genelec 8330A SAM<sup>28</sup> missati in stereo (L + R), mentre per la performance pubblica sono stati utilizzati quattro Genelec 1030A<sup>29</sup> in dotazione al Conservatorio, missati in doppio stereo (LL + RR). Si è evitato l'utilizzo di un woofer aggiuntivo considerando che il range d'altezza delle note in *Vexations* si muove tra Solb3 e La5 (con La4 = 440 Hz, quindi tra 196 Hz e 880 Hz).

### 2.3.4 Computer e schermo secondario

Il computer utilizzato sia per le prove che per la performance pubblica è stato un MacBook Pro Retina della metà del 2015<sup>30</sup>.

Uno schermo secondario da 60 pollici<sup>31</sup> in dotazione al Conservatorio è stato collegato al computer e utilizzato per il feedback visivo destinato al pubblico.

### 2.3.5 Piattaforma interattiva

Per "piattaforma interattiva" s'intende in questo caso un software basato su Max 7 di Cycling '74<sup>32</sup>, programmato per monitorare e mettere in relazione i dati dal corpo e i suoni dal pianoforte. L'interfaccia è stata concepita sia come strumento operativo per il performer, che come feedback visivo per il pubblico, e si divide in settori, alcuni dei quali destinati unicamente al performer (per il funzionamento della piattaforma, la registrazione dei dati rilevati dai sensori durante l'esecuzione<sup>33</sup>, l'equalizzazione degli altoparlanti in base alla risposta acustica dell'ambiente che ospita la performance), e altri – contenenti elementi grafici dinamici – destinati anche al pubblico in quanto utili alla comprensione delle relazioni tra l'attività sensoristica sul corpo e l'elaborazione del suono che ne deriva: il *Counter* (che visualizza la durata corrente della performance in giorni, ore, minuti, secondi e numero di ripetizioni del *motif* trascorse), il settore dei tre sensori e-Health/Arduino con le loro relative funzioni

<sup>27</sup> Esempi audio sono presenti nel documentario di A. Castaldini *Vexations: Soliloquy* (2017) di cui è disponibile un estratto alla pagina web <https://mnt.conts.it/production.php?page=ricerca>

<sup>28</sup> 104 dB max., woofer 50 W + tweeter 50 W, 45–23000 Hz.

<sup>29</sup> 115 dB max., woofer 80 W + tweeter 50 W, 52–20000 Hz.

<sup>30</sup> Processore 2,8 GHz Intel Core i7, memoria 16 GB, 1600 Mhz DDR3, grafica Intel Iris Pro 1536 (VRAM dinamica: 1536 MB) + AMD Radeon R9 M370X (VRAM dedicata: 2048 MB).

<sup>31</sup> TV LCD Sharp LE652.

<sup>32</sup> <https://cycling74.com>

<sup>33</sup> Al termine della performance, i dati raccolti vengono automaticamente visualizzati in forma testuale, per essere salvati come fogli di calcolo utili per analisi successive, oppure per essere caricati nuovamente nella piattaforma a replicare esattamente i cambiamenti psicofisici avvenuti in esecuzione.

<sup>22</sup> <http://opensoundcontrol.org>

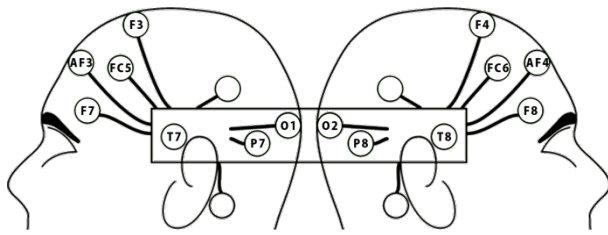
<sup>23</sup> In ambito neurologico, le bande di frequenza sono intervalli nel dominio delle onde cerebrali.

<sup>24</sup> L'unità è comprensiva dei sensori, e le librerie per la programmazione in Arduino IDE sono scaricabili dal sito del produttore.

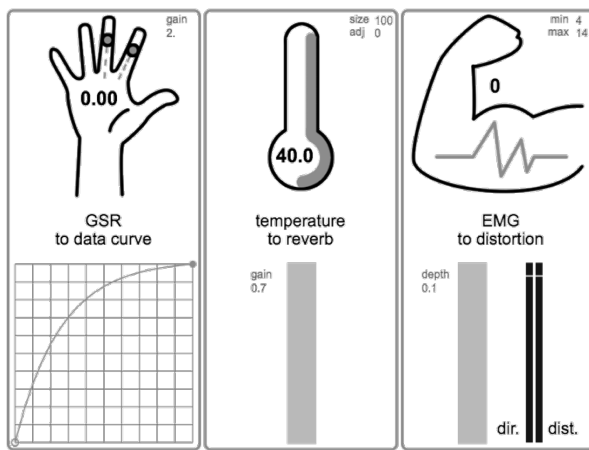
<sup>25</sup> <https://www.cooking-hacks.com>

<sup>26</sup> <https://www.arduino.cc>

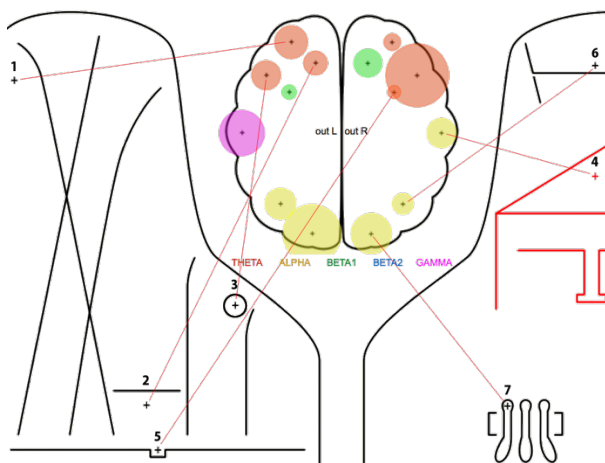
(*GSR to data curve*, *Temperature to reverb*, e *EMG to distortion*, Figura 3), e il settore dei sensori EEG con i collegamenti cervello–pianoforte (Figura 4).



**Figura 2.** Sigle dei sensori EEG dell'EPOC+ secondo il Sistema 10–20.



**Figura 3.** *GSR to data curve*, *temperature to reverb*, *EMG to distortion*.



**Figura 4.** Collegamenti cervello–pianoforte.

Quest'ultimo settore rappresenta il fulcro della piattaforma in quanto si occupa del collegamento tra regioni corticali del performer e zone timbriche del pianoforte, e pertanto costituisce la base logica e concettuale della performance. Nel feedback visivo sono riconoscibili le connessioni tra i 14 elettrodi encefalici e i 7 microfoni, dove ogni microfono è collegato a una coppia di elettrodi in posizione simmetrica:

1. 1° microfono side–address collegato a AF3/AF4;
2. 2° microfono side–address collegato a F3/F4;

3. 1° microfono a clip collegato a F7/F8;
4. 2° microfono a clip collegato a T7/T8;
5. 1° microfono a contatto collegato a FC5/FC6;
6. 2° microfono a contatto collegato a P7/P8;
7. 3° microfono a contatto collegato a O1/O2.

I sensori nell'emisfero destro sono coinvolti nell'elaborazione dei segnali microfoniche indirizzata al canale destro dell'uscita stereo, mentre i sensori di sinistra sono coinvolti nell'elaborazione dei segnali provenienti pur dagli stessi microfoni, ma destinata questa volta al canale sinistro dell'uscita stereo. Sempre nel feedback visivo, i 14 elettrodi sono graficamente rappresentati da cerchi colorati a diametro variabile. I colori rappresentano le diverse bande di frequenza cerebrale predominanti rilevate dai sensori in un dato momento: rosso/Theta, giallo/Alpha, verde/Low–Beta, blu/High–Beta, viola/Gamma. La piattaforma è programmata per far corrispondere ognuna delle cinque bande di frequenza (ognuno dei cinque colori) a una diversa regione dello spettro audio, con FFT a 256 campioni:

- Theta = campioni 0–4;
- Alpha = campioni 4–7;
- Low–Beta = campioni 7–11;
- High–Beta = campioni 11–18;
- Gamma = campioni 18–250.

A seconda della banda di frequenza cerebrale predominante rilevata da un sensore, il software ritaglia dall'audio del microfono associato a quel sensore solo la porzione di campioni FFT corrispondente alla banda rilevata: quindi dai 7 microfoni collegati ai 7 sensori dell'emisfero destro si ritagliano contemporaneamente 7 porzioni che vengono missate nel canale destro; allo stesso tempo, dagli stessi 7 microfoni collegati però ai 7 sensori dell'emisfero sinistro si ritagliano altre 7 porzioni che vengono missate nel canale sinistro. Ogni porzione ha una sua ampiezza indipendente che deriva dall'ampiezza della banda di frequenza predominante rilevata da ciascun sensore (rappresentata graficamente dall'ampiezza dei cerchi colorati nel feedback visivo). Ovviamente, la scala d'ampiezza delle onde cerebrali non combacia linearmente con la scala d'ampiezza audio: una corrispondenza musicalmente efficace è data da una curva, che in questo caso non è statica, ma viene fatta dipendere dinamicamente dalla risposta galvanica della pelle del performer. Il dato della risposta galvanica (rilevato dall'unità e–Health/Arduino ed espresso in volt) viene riscaldato nel sotto–settore *GSR to data curve*, ed utilizzato per impostare una curva che modella il range dinamico dei suoni in entrata. Gli altri due sotto–settori dell'unità e–Health/Arduino concorrono a caratterizzare il timbro del suono in uscita: *Temperature to reverb* utilizza il dato di temperatura corporea<sup>34</sup> del performer per controllare in modo

<sup>34</sup> Da numerosi test empirici di confronto tra la temperatura rilevata dal sensore e–Health e quella rilevata da vari termometri a mercurio, si è notata una sostanziale imprecisione da parte del sensore e–Health nell'individuazione del primo dato di temperatura, e una maggiore precisione nel rilevare le variazioni di temperatura nel tempo. Per

sottrattivo<sup>35</sup> la quantità di un effetto riverbero (più la temperatura scende, più il riverbero digitale aumenta), mentre *EMG to distortion* utilizza il dato di tensione muscolare<sup>36</sup> per controllare additivamente<sup>37</sup> la quantità di un effetto distorsivo di tipo chorus (più la tensione muscolare aumenta, più il suono in uscita viene distorto). Pertanto, la crescente difficoltà fisica del performer nell'arco dell'esecuzione (il calo di temperatura, la crescente tensione muscolare) è resa udibile dalla graduale distorsione e perdita di nettezza del "profilo" sonoro del pianoforte. Questo nesso e la precedente scelta dei collegamenti tra sensori e microfoni – quindi tra regioni corticali e zone timbriche dello strumento – sono certamente arbitrari, dipendono cioè dal gusto del performer, ma si basano sull'attenta osservazione dell'attività neuronale durante i test domestici. Ad esempio, dopo aver osservato che la regione frontale del cervello ospitava frequentemente onde Theta in relazione agli stati di maggior "reattività" emotiva, di agitazione o ansia, si è deciso di legare le coppie di sensori AF3/AF4, F3/F4 e F7/F8 ai tre microfoni destinati alle zone del pianoforte maggiormente risonanti e acusticamente più "piene" nella parte bassa dello spettro audio, ovvero i due side-address e il 1° a clip. Mentre le zone parietale (P7/P8) e occipitale (O1/O2), manifestando più spesso onde Alpha associate soprattutto a momenti di rilassamento e agli occhi chiusi, sono state collegate al 2° e al 3° microfono a contatto destinati alle zone timbriche più "argente" del pianoforte (ricche cioè nella parte alta dello spettro sonoro). Si può quindi dire che a differenza del meccanismo di missaggio – dove i colori dei cerchi rappresentano la relazione diretta tra l'altezza delle bande di frequenza predominanti e l'altezza dei ritagli di spettro audio corrispondenti – l'associazione tra regioni corticali e zone timbriche è piuttosto "a programma", ovvero attribuisce una sorta di significato musicalmente "espressivo" alla maggior presenza di alcune bande predominanti in determinate regioni corticali. In quest'ottica, forse meno convincenti sono risultate le combinazioni FC5/FC6 con il 1° microfono a contatto attaccato alla serratura del piano, e T7/T8 con il mi-

---

questa ragione è stato introdotto nella piattaforma un sistema di taratura manuale con cui impostare il primo dato di temperatura corporea prima dell'inizio della performance.

<sup>35</sup> Il range 30–40 °C viene riscaldato, in modo lineare-inverso, nei valori decimali 0.9–0.05.

<sup>36</sup> Nel caso specifico della performance del 2016 è stato scelto il muscolo bicipite brachiale.

<sup>37</sup> Il dato di attività elettromiografica espresso in volt viene riscaldato linearmente in valori decimali (0.0–1.0). Tuttavia il dato in volt viene ritardato automaticamente ogni 100 secondi. Infatti, nel corso dei test domestici è stata notata una perdita di taratura nel tempo da parte del sensore a gel (unico sensore a gel tra tutti i sensori utilizzati): non appena il gel contenuto nell'elettrodo si secca (oltre una durata di circa 4 o 5 ore) i valori minimo e massimo restituiti dal sensore traslano repentinamente di una medesima ma apparentemente casuale quantità. Per questa ragione, la piattaforma è stata programmata per ritardare automaticamente i valori minimo e massimo ogni 100 secondi riportandoli al range iniziale 4–14 (range tarato sperimentalmente sul performer). La finestra di 100 secondi serve a coprire abbondantemente la durata di un *motif* completo.

crofono a clip schiacciato contro il coperchio chiuso: la varietà delle bande di frequenza osservata nelle regioni temporali – con presenza anche di predominanti Beta e Gamma, molto rare nelle altre regioni corticali – è stata inizialmente risolta collegando T7/T8 alla zona spettralmente più uniforme del pianoforte – tra il centro della cassa armonica e il coperchio – affinché qualsiasi banda predominante potesse contare sulla presenza di energia sonora nella corrispondente porzione di spettro audio da ritagliare. In seguito si è però osservato che le onde Beta e Gamma erano molto più visibili nei momenti in cui il performer era impegnato in un'attività specifica del pensiero cognitivo convergente (come il calcolo a mente) o nel problem solving (ad esempio, nel risolvere imprevisti ambientali come la presenza di forti rumori dalla strada che imponevano una correzione manuale dell'equalizzazione), oppure ancora in un'attività diagnostica (come monitorare il corretto funzionamento della piattaforma interattiva e la distribuzione del carico di lavoro tra i core del processore). Probabilmente questa attività mentale sarebbe stata musicalmente meglio rappresentata dal timbro puntillistico ottenuto dal microfono posto sulla serratura del pianoforte.

### 2.3.6 Superfici di controllo MIDI

La piattaforma offre la possibilità di agire su tutti i controlli dell'interfaccia per trascinarsi del mouse sui campi numerici, oppure tramite superfici esterne di controllo MIDI. La piattaforma prevede due controller slider e un controller a pulsanti. Nella performance del 2016 sono stati utilizzati i controller slider iControls (ICON) e nanoKONTROL2 (Korg), e il controller a pulsanti MIDI Fighter 3D (DJ TechTools). La scelta del MIDI Fighter 3D è stata motivata dalla quantità e tipologia dei pulsanti sulla superficie di controllo: tutti i 16 leggerissimi pulsanti Sanwa<sup>38</sup> sono stati programmati per registrare sul *Counter* il numero corrente di ripetizioni del *motif* durante la performance (ogni volta che il *motif* veniva ripetuto, il performer doveva premere uno dei 16 pulsanti del controller affinché il *Counter* registrasse l'avvenuta ripetizione). La scelta di programmare tutti i tasti del controller per la medesima funzione si spiega per la difficoltà di effettuare una pur elementare operazione 840 volte nell'arco di 24 ore: un eventuale stato di trance o anche solo un banale calo attentivo avrebbero potuto portare il performer a sbagliare pulsante, o a non premerlo fino in fondo, mancando così la registrazione del *Counter*. Destinando invece tutti i pulsanti al medesimo scopo, era sufficiente far cadere la mano aperta sul controller per essere sicuri di non commettere errori<sup>39</sup>.

---

<sup>38</sup> I pulsanti Sanwa sono dello stesso tipo di quelli utilizzati per i videogiochi arcade: larghi – ø 24mm – e leggerissimi, pensati per un utilizzo intensivo.

<sup>39</sup> Per evitare però che il *Counter* registrasse molti impulsi a causa dell'eventuale pressione simultanea di più pulsanti, la piattaforma è stata programmata per accettare una sola pressione ogni 60 secondi, ovvero una durata maggiore della metà del *motif* – momento in cui il performer suona con una mano sola la ripetizione del *Thème*, e po-

## 2.4 Training psicofisico

La preparazione alla performance del 2016 ha richiesto un training psicofisico accelerato e una dieta specifica per circa due mesi e mezzo, sotto la guida di un'equipe di Medicina Sportiva dell'Ospedale di Noale (Venezia) diretta dalla dott.ssa Erica Brugin<sup>40</sup>.

## 2.5 Location e forma d'ascolto

La location della performance del 2016 è stata Piazza dell'Unità (Trieste) all'aperto. Lo spazio esecutivo è stato simbolicamente arginato da un gazebo di 9 m<sup>2</sup> (Figura 5) che aveva la funzione di proteggere le apparecchiature dagli agenti atmosferici, ma era aperto ai lati. La scelta di collocare l'esecuzione in uno spazio all'aperto è stata coerente con la volontà di massimizzare gli stimoli ambientali e ottenere reazioni

trebbe erroneamente credere di star invece eseguendo il *Thème* all'inizio del *motif*, e di dover quindi premere il pulsante di conteggio. Se questo fosse avvenuto, grazie al blocco di 60 secondi, il *Counter* non avrebbe registrato l'avvenuta ripetizione.

<sup>40</sup> Si riporta il report della dott.ssa Brugin: «Il soggetto è stato sottoposto ad un'iniziale valutazione medico-sportiva con la dott.ssa Erica Brugin, che ha compreso anamnesi completa, visita clinica, esami bioumorali, ECG a riposo e test da sforzo cardiopolmonare massimale al cicloergometro. Il test ha evidenziato una ridotta capacità aerobica, per cui è stato impostato un allenamento personalizzato consistente in almeno 3 sedute settimanali di esercizio aerobico della durata di almeno 45 minuti, e di intensità moderata prescritta indicando il range appropriato di frequenza cardiaca (calcolato in base ai risultati del test da sforzo).

Il soggetto è stato inoltre sottoposto ad esame posturale con il dott. Luigi Gerbino, e ad una serie di test sottomassimali per la valutazione della forza dei principali gruppi muscolari corporei con il dott. Ruggero Marin. L'analisi ha evidenziato una forza muscolare inferiore al predetto per genere ed età, con ipostenia in particolare a carico dei muscoli stabilizzatori del tronco e dei muscoli del cingolo scapolare. Sono stati prescritti specifici esercizi di rinforzo muscolare, in particolare a carico dei distretti più deboli, associando anche esercizi volti al miglioramento della coordinazione motoria. Il training di rinforzo muscolare ha previsto alcune sedute di somministrazione supervisionata per un corretto apprendimento dei movimenti, ed è stato poi continuato a domicilio dal soggetto in autonomia.

Le dott.sse Luisa Bedin e Giovanna Tegon hanno valutato l'aspetto nutrizionale del soggetto attraverso l'analisi delle abitudini alimentari (tramite diario domiciliare settimanale) e l'esecuzione di una bioimpedenziometria per analizzare la composizione corporea. Sono stati forniti consigli alimentari personalizzati basati in particolare su un aumento dell'idratazione, una corretta distribuzione circadiana degli alimenti ed un'appropriata scelta quantitativa e qualitativa degli stessi indicando in particolare l'aumento del contenuto di fibre alimentari e proteine nella dieta, e la riduzione di zuccheri semplici. È stato inoltre indicato l'uso di integratori specifici (sempre in base al risultato della valutazione impedenziometrica) tenendo in considerazione gli elevati valori di ferritina presentati dal soggetto.

Il paziente ha partecipato a 8 incontri psicologici della durata di un'ora a cadenza settimanale con la dott.ssa Marzia Sarto. Partendo da una valutazione della struttura di personalità, è stato impostato un programma volto a limitare le ansie soggettive attraverso un percorso breve di ristrutturazione cognitiva ove necessario. Tecniche utilizzate: tecnica di rilassamento di Jacobson e Metal Training (per favorire la congiunzione tra visualizzazione dell'immagine della prestazione svolta al meglio e la gestione delle emozioni reattive all'evento stesso). Il paziente si è allenato a domicilio con successiva verifica in studio dei risultati ottenuti per adeguare e personalizzare le tappe successive.

È stata curata la regolarizzazione del ritmo sonno veglia anche mediante utilizzo di melatonina.

Sono state, infine, pensate soluzioni pratiche per le necessità fisiologiche richieste dalla performance prolungata (quali, ad esempio, ausili assorbenti per l'attività urinaria).»

psicofisiche significative da parte del performer: le variazioni di temperatura atmosferica e di luce naturale, e la vicinanza e libertà comportamentale dei visitatori/spettatori e semplici passanti sono stati la spinta propulsiva per l'elaborazione del suono.



**Figura 5.** Il gazebo in Piazza dell'Unità prima della performance (foto: © 2016 Roberto Srelz).

## 2.6 Esecuzione

Prima dell'esecuzione, il pianista ha svolto alcuni esercizi di rilassamento e mental training.

L'esecuzione è poi iniziata ed è proceduta senza intoppi: sono state rispettate le tempistiche relative al cibo<sup>41</sup> e all'umettatura dei sensori EEG con soluzione salina<sup>42</sup>.

Durante tutta la performance non si sono verificati fenomeni piovosi, e l'escursione termica è stata contenuta, pertanto gli elementi atmosferici non hanno esercitato variazioni significative sullo stato fisico del pianista.

Già dopo la sua prima esposizione, il *motif* è stato assimilato dalla memoria "meccanica" (ovvero la memoria di movimento delle dita, grazie soprattutto all'utilizzo di una speciale diteggiatura). Il pianista ha quindi potuto proseguire senza dover mantenere un'attenzione costante durante tutta l'esecuzione.

## 2.7 Dialogo con il pubblico

Nonostante all'esterno del gazebo fosse stato affisso un cartellone bilingue (italiano e inglese) con una descrizione del concept alla base della performance e una sommaria spiegazione del sistema interattivo utilizzato, durante l'esecuzione molti spettatori si sono spontaneamente avvicinati al performer per interloquire. Le domande vertevano principalmente sui dettagli tecnici del sistema sensoristico e sul metodo utilizzato

<sup>41</sup> Il pianista è stato imboccato da un assistente ogni 3 ore con pezzi di frutta ed integratori, ed idratato con acqua, secondo la prescrizione dell'equipe di Medicina Sportiva di Noale.

<sup>42</sup> L'umettatura dei sensori EEG è avvenuta ogni ora circa, sia per mantenere attivo il contatto degli elettrodi sul cuoio capelluto, che per lenire l'irritazione cutanea dovuta alla pressione continua dell'estensione di feltro semirigido dei sensori.

per convertire i dati provenienti dal corpo in valori di controllo del suono. Si può probabilmente spiegare questa curiosità tecnica considerando le aspettative del pubblico che in quei giorni frequentava il festival scientifico–tecnologico Trieste Next, e che considerava la performance uno degli eventi di carattere scientifico proposti. Il pubblico, vario per età, genere e livello di formazione (tra chi ha posto domande c'erano anche bambini), ha generalmente espresso una certa sorpresa nel veder utilizzare sistemi di ambito medico–scientifico in un'esecuzione musicale pianistica, e soddisfazione nella comprensione dei meccanismi messi in gioco per la conversione dei dati corporei e cerebrali in valori di controllo del suono, giudicando la performance un "progresso" nella pratica esecutiva strumentale. Tuttavia, dalle domande rivolte al performer, non si è notata altrettanta curiosità per gli aspetti più squisitamente musicali ed estetici o per il "senso" artistico del progetto.

### 2.8 Controllo e simulazione degli stati emotivi

L'interlocazione diretta con il pubblico è stato un fatto non previsto e inusuale in un'esecuzione pianistica, e uno stimolo molto forte per il performer: la difficoltà di eseguire *Vexations* al pianoforte e allo stesso tempo interloquire col pubblico ha inizialmente prodotto una dominanza di onde Theta con forti picchi d'ampiezza, che in virtù dell'elaborazione audio hanno comportato episodici effetti Larsen. Il performer ha quindi messo in atto un tentativo di controllo dello stato emotivo che, grazie alle tecniche di rilassamento acquisite all'ospedale di Noale, è stato raggiunto e mantenuto con sempre maggior efficacia nell'arco dell'intera esecuzione. Nelle ultime ore, il performer è arrivato a pilotare consciamente gli effetti audio simulando diversi stati emotivi (grazie soprattutto all'exasperazione delle espressioni facciali mimiche<sup>43</sup>). Questa possibilità – ovvero la gestione cosciente degli effetti audio attraverso la simulazione degli stati emotivi – non era stata contemplata nel progetto originario, ma è stata comunque accolta favorevolmente dal performer e dagli spettatori. Il "colloquio" con il pubblico è infatti avvenuto non solo sul piano strettamente verbale, ma anche nel tentativo da parte del performer di controllare "a richiesta" gli effetti audio, allargando così l'idea originaria di una performance condizionata dagli stimoli ambientali.

### 3. CONCLUSIONI

L'ipotesi alla base di questa ricerca, ovvero che l'annotazione di Satie avvicini *Vexations* ad un progetto di body art (sez. 2.1) appare sostanzialmente

confermabile (la prudenza è d'obbligo data l'intrinseca soggettività di qualsiasi ricerca di natura artistica). Il brano di Satie si è inserito perfettamente nella performance scientifico–musicale di cui era l'origine ispirativa, assimilando i dispositivi sensoristici e le aumentazioni audio del pianoforte come naturali estensioni dei mezzi d'interpretazione musicale (non già come mera aumentazione elettronica per un arrangiamento "tech" votato al pretenzioso svecchiamento di un brano di repertorio). Senza aver apportato alcuna modifica alla partitura originale e rispettando le indicazioni dell'autore, l'interpretazione nella sua aumentazione ha comportato la definizione di uno specifico spazio fisico d'ascolto e una coerente modalità di fruizione musicale per il pubblico.

Resta tuttavia aperta la questione della qualità musicale della performance: se da un punto di vista performativo, la lettura body–artistica di *Vexations* ha effettivamente dimostrato una naturale efficacia (considerando anche il modo in cui il pubblico ha spontaneamente riconosciuto e approcciato l'opera, ovvero con gli stessi tempi e modi con cui solitamente si appropria un'installazione artistica, non già un recital pianistico), da un punto di vista invece più specificamente musicale, l'effettivo risultato sonoro (il flusso ininterrotto del "morphing" audio dovuto alla continua trasmissione di dati EEG alla piattaforma interattiva) è risultato sovrastato in importanza dal concept. In altre parole, al di là di essere stata percepita come segnale del mutamento psicofisico del pianista, l'elaborazione del suono è apparsa non determinante nella sua qualità musicale. È doveroso quindi pensare ad un miglioramento degli effetti audio e ad un diverso mapping tra dati sensoristici e valori di controllo del suono, e forse anche alla possibilità di soppiantare il "morphing" audio ininterrotto con un sistema più sofisticato di "dosatura" degli effetti in archi temporali prestabiliti.

### 4. REFERENCES

- [1] E. Satie: "Vexations" *Contrepoints, une revue de musique*, No. 6, 1949.
- [2] E. Satie: "Vexations", *Art News Annual*, Vol. 27, 1958.
- [3] G. Bryars, "Vexations and its Performers", *Contact: A Journal of contemporary Music*, No. 26, p. 12, 1983.
- [4] C. Kohlmetz, R. Kopiez, E. Altenmüller: "Stability of motor programs during a state of meditation: Electrocortical activity in a pianist playing 'Vexations' by Erik Satie continuously for 28 hours", *Psychology of Music*, No. 31(2), pp. 173–186, 2003.

<sup>43</sup> Secondo la documentazione relativa al casco EPOC+ (reperibile nel sito [emotiv.zendesck.com](http://emotiv.zendesck.com)), la rilevazione delle *Facial Expressions* risulta di norma più evidente (in termini di potenza captata dai sensori) rispetto ai *Performance Metrics* (ovvero gli autentici stati emotivi) che quindi possono essere mascherati.

# Applicazioni storiche della Teoria del Caos nella computer music

**Luca Morino**

Conservatorio “G. Verdi” di Torino  
morinoluc@gmail.com

**Daniele Ghisi**

Conservatorio “N. Paganini” di Genova  
danieleghisi@bachproject.net

## ABSTRACT

Questo articolo presenta una panoramica storica sulle implicazioni teoriche e sulle applicazioni pratiche che la Teoria del Caos ha determinato nel pensiero e nei processi creativi di alcuni compositori d'avanguardia di fine Novecento, in particolare nell'ambito della *computer music*, esponendo alcuni concetti fondamentali del *caos deterministico* e proponendo una selezione di tecniche compositive basate su algoritmi caotici.

## 1. INTRODUZIONE

La Teoria del Caos è stata considerata da molti [1] la terza grande rivoluzione del pensiero scientifico del XX secolo, dopo la Teoria della Relatività e la Teoria Quantistica. Essa, infatti, rivoluzionò l'approccio che fino ad allora la comunità scientifica aveva avuto nell'osservazione del mondo, riportando in primo piano alcuni fenomeni che erano stati in parte ignorati o poco approfonditi, in parte ritenuti di scarso interesse. Fino ad allora si era ricercato nei fenomeni naturali un ordine, una linearità, un'armonia che, però, non sono proprie del mondo reale, se non con le dovute approssimazioni. Queste proprietà sono, in realtà, le più facili da osservare e da analizzare in un fenomeno fisico. Per lungo tempo, infatti, si conservarono l'ottimismo e la mentalità del positivismo ottocentesco. Per dirla con le parole di Laplace:

“Nous devons envisager l'état present de l'univers comme l'effet de son état antérieur et comme la cause de celui qui va suivre. Une intelligence qui, pour un instant donné, connaîtrait toutes les forces dont la nature est animée et la situation respective des êtres qui la composent, si d'ailleurs elle était assez vaste pour soumettre ces données à l'Analyse, embrasserait dans la même formule les mouvements des plus grands corps de l'univers et ceux du plus léger atome; rien ne serait incertain pour elle, et l'avenir comme le passé serait present à ses yeux [2].”<sup>1</sup>

<sup>1</sup> “Noi dobbiamo considerare lo stato presente dell'universo come l'evoluzione del suo stato passato, e come la causa di quello che accadrà. Un'intelligenza che, ad un certo istante, potesse conoscere tutte le forze

Durante il Novecento, alcune rivoluzioni nel campo della matematica, come, ad esempio, il Teorema di Incompletezza di Gödel [3], avevano minato l'atmosfera di onnipotenza che aleggiava intorno al pensiero scientifico. I risultati dimostrati dalla Teoria del Caos smantellarono una volta per tutte l'idea di un mondo meccanico perfettamente prevedibile per quanto complesso esso sia, come ipotizzato da Laplace. La realtà che ci circonda è un sottile e instabile equilibrio di una struttura diversa, polimorfa [4]. La comunità scientifica dovette quindi prendere coscienza che anche un sistema formalizzato e totalmente deterministico può assumere dei comportamenti intrinsecamente non prevedibili.

Si iniziò a parlare di *caos deterministico* a partire dalla prima metà degli anni Settanta, in seguito alla scoperta e allo studio di particolari sistemi che, senza alcun elemento aleatorio o stocastico al loro interno, manifestano un comportamento caotico.

Questi sistemi sono di varia natura. La Teoria del Caos, così compatta e potente nel suo valore e significato teorico, trova il suo massimo centro di applicazione nei *sistemi dinamici non-lineari* [5], arrivando a comprendere anche altre strutture con caratteristiche peculiari come, ad esempio, gli *L-Systems*, i *frattali* [6] e gli *automi cellulari* [7].

Tali strutture sono intimamente legate e accomunate, tra le altre cose, dai concetti di *retroazione*, o *feedback*, e di evoluzione autonoma e autoalimentata nel tempo o nello spazio di un particolare sistema.

A partire dalle avanguardie novecentesche, la casualità è diventata uno strumento compositivo di rilievo. Dai lavori di Cage, Ligeti o Xenakis, che hanno fatto ampio uso di algoritmi stocastici nei loro processi creativi, molti compositori contemporanei hanno indagato il comportamento dei sistemi caotici applicati a varie tecniche di composizione musicale, in particolare nell'ambito della *computer music*. Il caos deterministico ebbe quindi molta fortuna tra coloro che si cimentarono nella composizione assistita al computer e nella sintesi digitale del suono, in particolare durante gli anni Novanta.

Questo articolo, estratto dal lavoro *Caos deterministico: la Teoria del Caos nella computer music* [8], presenta una ricerca sulle implicazioni che il caos deterministico ebbe nell'ambito della *computer music*.

che animano la natura, e la posizione relativa di tutto ciò che la compone, e che fosse anche così vasta da riuscire a sottoporre questi dati all'Analisi, potrebbe condensare in una sola formula i movimenti dei corpi più grandi dell'universo e quelli dell'atomo più leggero; nulla per lei sarebbe incerto, e ai suoi occhi sarebbero presenti sia l'avvenire che il passato.”, traduzione degli autori, tratto da *Théorie analytique des probabilités*, di Pierre Simon de Laplace.

Copyright: ©2018 Luca Morino et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution License 3.0 Unported](https://creativecommons.org/licenses/by/3.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

## 2. CAOS E FORMALIZZAZIONE ALGORITMICA NELLA COMPUTER MUSIC

Gli algoritmi sono la linfa vitale della *computer music*: a loro viene spesso delegato il compito di regolare vari aspetti della composizione musicale, dalla creazione e controllo dello spettro sonoro alla generazione di strutture musicali.

### 2.1 Dal caso al caos

Gli algoritmi stocastici sono per natura intrinsecamente dipendenti dalle leggi di probabilità e rendono teoricamente impossibile prevedere l'esito esatto del processo in qualsiasi punto del futuro [9]. Storicamente hanno da sempre avuto ampio spazio nella *computer music*. Ad esempio, il moto browniano (anche detto *random walk*), le *catene markoviane* e le varie forme di *fractional noise* sono ancora oggi tecniche abbondantemente utilizzate nei processi compositivi [10].<sup>2</sup> *The Illiac Suite for String Quartet* di Lejaren Hiller (1957), considerato il primo brano di *computer music*, poiché la partitura del quartetto è stata completamente generata da un computer, era in buona parte basato su algoritmi aleatori.

Occorre sottolineare che l'idea di caos era già presente in alcuni grandi compositori prima dell'era cybernetica. Uno dei primi compositori che fece della casualità assoluta un paradigma compositivo è John Cage. I processi aleatori che stanno alla base di molti suoi lavori sono utilizzati per il loro forte valore concettuale, più che formale. Infatti, la poetica di Cage si basa spesso sul concetto di contingenza degli eventi: il caso non solo concorre allo sviluppo dell'estetica cageana, ma diventa la musica stessa.

Un altro grande compositore che si avvicinò a processi stocastici perché insofferente al serialismo esasperato degli anni Cinquanta fu Iannis Xenakis. Il suo impiego di processi stocastici è, però, fortemente diverso da quello di Cage: egli applicava processi aleatori vincolati da un numero variabile di condizioni predeterminate. Se per Cage l'applicazione del caso a livello compositivo era un atto di emancipazione della musica, affinché si svincolasse in primis dall'autore, ma anche da altri fattori prestabiliti che potesse condizionarla, per Xenakis la casualità nell'atto compositivo era un via per mettere a nudo e validare il processo formale compositivo a più alto livello. Un brano significativo in tal senso è *Analogique A+B*, per nove archi ed elettronica (si veda la Fig. 1). Scritto nel 1959, è un brano in cui l'elettronica diffusa in ottofonìa e la piccola orchestra d'archi si alternano e si fondono continuamente. Oltre ad essere uno dei primi esempi di composizione algoritmica - la partitura degli archi è stata generata con

<sup>2</sup> Il moto browniano è un processo iterativo stocastico con un grado di memoria, ovvero il valore al passo successivo è determinato casualmente in un intorno però definito dal valore precedente, quindi vincolato alla evoluzione del sistema. Anche le catene markoviane sono procedimenti stocastici con uno o più gradi di memoria. I vincoli vengono però calcolati su uno o più valori precedenti a partire da una matrice di analisi (matrice di trasferimento) preformata. Fractional noise è invece un metodo di generazione stocastico vincolato da determinate relazioni definite a livello frequenziale. È, per esempio, utilizzato per sintetizzare varie forme di rumore: il White Noise (rumore bianco) ha massimo grado di casualità, definito dalla relazione spettrale  $1/f^0$ . Il Pink Noise (rumore rosa) ha invece densità spettrale definita  $1/f^1$ . Il Brown Noise (rumore marrone) ha un vincolo più forte: spettralmente è definito dalla relazione  $1/f^2$ . È inoltre analogo al moto browniano [11].



Figure 1. Iannis Xenakis, partitura di Analogique A (1958, 9 strings): battute 1-3.

processi algoritmi stocastici (*Analogique A*) - questo brano è anche conosciuto per essere uno dei primi in cui si utilizzano tecniche per la sintesi granulare (*Analogique B*).

Se considerassimo il caos deterministico come “evoluzione concettuale” degli algoritmi stocastici, potrebbe essere interessante notare come l'utilizzo del caso nelle pratiche compositive abbia avuto la stessa evoluzione consequenziale: dalla casualità quasi assoluta e non mediata alla strutturazione di processi formali sempre più definiti entro cui contenerla.

La teorizzazione del caos deterministico è relativamente recente ed è fortemente legata allo sviluppo delle capacità di calcolo dei computer: a partire dagli anni Ottanta la sua divulgazione suscitò l'interesse di giovani compositori-ricercatori nell'ambito della *computer music*. In molti videro nel comportamento peculiare dei sistemi dinamici non-lineari ed altri algoritmi caotici una nuova possibilità espressiva, un punto d'interpolazione tra un processo stocastico e un processo deterministico. L'interesse fu vivo, vennero pubblicati molti studi a riguardo e le aspettative erano molte: “...new music models will undoubtedly arise from the intellectual milieu that includes fractal geometry and chaotic non-linear systems [12].”<sup>3</sup>

### 2.2 Il “formalismo frattale”

Un altro vasto aspetto che concerne la Teoria del Caos, come detto, sono le geometrie *frattali*. Il concetto fondamentale, in questo caso, è l'autosimilarità. Concetto che non è certamente nuovo nella prassi compositiva: già in Bach troviamo un esempio di figure che si ripropongono a diversi livelli. Le tecniche compositive di “accrescimento” o “diminuzione”, ad esempio, permettono di creare strutture autosimili a livello ritmico o melodico. Tuttavia, pochi sono i casi precedenti al Novecento in cui il principio

<sup>3</sup> “Indubbiamente sorgeranno nuovi modelli musicali da un ambiente intellettuale che comprende la geometria dei frattali e i sistemi caotici non lineari”, traduzione degli autori.

di autosimilarità è applicato in modo formale e soprattutto strutturale.

Gli algoritmi iterativi che generano autosimilarità ebbero invece una vasta applicazione nei processi compositivi a partire dallo sviluppo delle tecniche di composizione seriale e, successivamente, nella *computer music*. Ad esempio, in alcune opere di Stockhausen, in cui viene applicato il concetto del continuum temporale tra *macrotempi* e *microtempi*<sup>4</sup> insieme ad un rigido serialismo formale, presentano strutture simili applicate a differenti livelli.

I primi esempi di “sonificazione” di frattali risalgono agli anni Ottanta. Nel 1981, Larry Austin compose il brano *Canadian Coastlines: Canonic Fractals for Musicians and Computer Band*, prima applicazione di una struttura frattale (la linea costiera del Canada). Commissionato dalla Canadian Broadcasting Corporation, *Canadian Coastline* è una composizione destinata alla trasmissione radiofonica in forma di canone a otto voci, quattro delle quali sono eseguite da otto musicisti, le restanti quattro dalla “*Computer Band*”, ovvero sequenze preregistrate di suoni sintetizzati al computer. Ogni voce esegue un’imitazione ritmico-melodica del tema, ma tutte le voci hanno un tempo metronomico diverso e mutevole. Un altro famoso esempio di sonificazione di frattali è il brano acusmatico *Profiles* di Charles Dodge [13], del 1988: uno dei primi brani di *computer music* in cui viene applicato in modo strutturale il principio di autosimilarità frattale dello spettro sonoro. La composizione, interamente elaborata al computer, si basa sull’applicazione sistematica del processo stocastico definito spettralmente dalla relazione  $1/f$  per la definizione di tutti i parametri musicali. Lo stesso Dodge, infatti, descrive il lavoro come un “frattale musicale”.

### 3. LE APPLICAZIONI MUSICALI DEL CAOS

Esistono sostanzialmente due approcci differenti tramite i quali i compositori hanno applicato gli algoritmi caotici nei loro processi creativi:

- le *applicazioni ad alto livello* si riferiscono al trattamento simbolico o numerico del materiale musicale, relativo ai parametri descrittivi di una nota o di un evento sonoro;
- le *applicazioni a basso livello* si concentrano invece sull’applicazione del caos a livello di generazione e trattamento del segnale, dalla sintesi del suono al controllo dei parametri spettrali.

In entrambi gli approcci, il modo più semplice di rappresentare musicalmente i sistemi caotici avviene mediante il *mapping* delle traiettorie tracciate nello spazio delle fasi di questi sistemi direttamente sui parametri musicali o del suono. Poiché il comportamento dei sistemi caotici può essere tanto regolare quanto imprevedibile ed eterogeneo, il mapping diretto sui vari parametri musicali può fornire una vasta gamma di risultati.

Partendo dal presupposto che un compositore che utilizzi un processo compositivo algoritmico decida di intervenire il meno possibile nella rielaborazione a posteriori

<sup>4</sup> Si pensi a *Kontakte*, 1960

dei valori ottenuti, l’impiego di sistemi caotici applicati alla composizione - come anche, spesso, nel caso di sistemi stocastici - comporta alcune problematiche. Una criticità fondamentale per la realizzazione di intenzioni specifiche compositivo-musicali dell’autore è la parvenza di mancanza di strutturazione musicale delegata all’algoritmo. Infatti, definito il sistema di equazioni e impostati i valori iniziali, il sistema produce nuovi eventi ad oltranza, la cui progressione è, di fatto, autonoma.

“...non vi è sviluppo in senso convenzionale - la procedura produce una certa naturale progressività che con metodi perfettamente deterministici (cioè esplicitamente finalizzati e guidati) sarebbe stata difficile ottenere. D’altra parte, la struttura del brano non è nemmeno aleatoria, poiché non presenterebbe altrimenti certe regolarità (ripetizioni correlate a medio e lungo termine) che diventano riferimenti percettivi essenziali nell’identificazione della forma del brano [14].”

Non è quindi possibile impostare a priori un comportamento formale minuto e puntiglioso, che rispecchi forme compositive convenzionali, ma, al più, un comportamento generale dell’evoluzione del sistema, il cui paradigma sarà legato alla quasi periodicità e caoticità degli algoritmi.

Naturalmente, il principio secondo cui l’applicazione algoritmica più “pura” di modelli fisico-matematici pone dei limiti al controllo del compositore sul risultato può manifestarsi nella maggior parte delle procedure di composizione algoritmica che implicino un sistema rigido di regole formali. La definizione di tali norme, di natura deterministica, permette tuttavia un controllo ad alto livello dei risultati musicali e la definizione di specifici movimenti.

È proprio il comportamento vario e complesso dei sistemi caotici deterministici che ha suscitato un forte interesse nella composizione algoritmica:

“The great attraction of nonlinear dynamical systems for compositional use is their natural affinity to the behaviors of phenomena in the real world, coupled with their mechanical efficiency of their computation and control. Chaotic systems offer a means of generating a variety of raw materials within a nonetheless globally consistent context. Chaotic sequences embody a process of transformation, the internal coherence of which is ensured by the rules encoded in the equations [15].”<sup>5</sup>

Nel momento in cui le intenzioni musicali del compositore assumono una concezione più astratta, in cui l’impiego dell’algoritmo e il controllo su di esso non sono finalizzate a riprodurre fedelmente un gesto musicale predefinito,

<sup>5</sup> “La grande attrazione che generano i sistemi dinamici non-lineari per un utilizzo compositivo sta nella loro naturale affinità con il comportamento dei fenomeni del mondo reale, unita all’efficienza di calcolo e al controllo del loro determinismo. I sistemi caotici offrono un mezzo per la generazione di una gran varietà di materiale in un contesto comunque globalmente coerente. Sequenze caotiche incorporano quindi un processo di trasformazione, la cui coerenza interna è assicurata dalle norme codificate nelle equazioni”, traduzione degli autori.

i sistemi caotici risultano essere potenti strumenti espressivi, perché richiedono il controllo di pochi parametri per ottenere risultati molto vari. In altre parole:

“I modelli di sistemi non-lineari permettono di gestire grandi quantità di dati in modo sostanzialmente diverso da quanto è possibile fare con altri mezzi. Nelle applicazioni che fanno uso di procedure di sintesi digitale del suono, la morfologia di interi eventi sonori è catturata nei soli valori di inizializzazione del sistema: il compositore può gestirne perfette riproduzioni, o alterazioni più o meno pronunciate, operando su due sole variabili [14].”

Tra le tante equazioni dinamiche non-lineari, sono quattro i modelli che hanno riscontrato più successo, forse perché i più intuitivi, forse perché mostrano sfaccettature peculiari del comportamento caotico: l'attrattore di Hénon (si veda Fig. 2), l'attrattore di Lorenz (si veda Fig. 3), l'attrattore di Duffing (si veda Fig. 4) e l'oscillatore di Chua (si veda paragrafo 3.3).

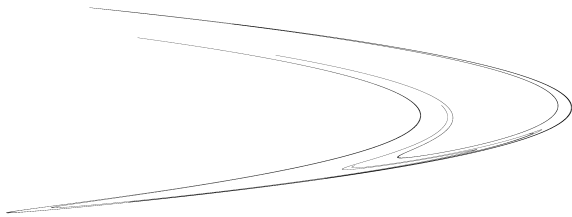


Figure 2. L'attrattore di Hénon

### 3.1 Applicazioni ad alto livello

Il primo approccio ai sistemi caotici sperimentato è stato quello ad alto livello, a cavallo tra gli anni Ottanta e Novanta. Pionieri in questo campo sono stati Jeff Pressing, Agostino Di Scipio, Michael Gogins e Richard Bidlack.

Nel 1988, Jeff Pressing pubblicò un articolo sul *Computer Music Journal* [16] in cui propose di applicare ai parametri musicali di una nota i valori ottenuti dallo sviluppo nel tempo di sistemi caotici a una, due, tre e quattro dimensioni, finalizzati alla generazione di una partitura per strumenti di sintesi sul programma Csound. Ne ricavò una quindicina di studi (*Sound Example n*), nei quali si concentrò sul confine che si delinea nello spazio delle fasi dei

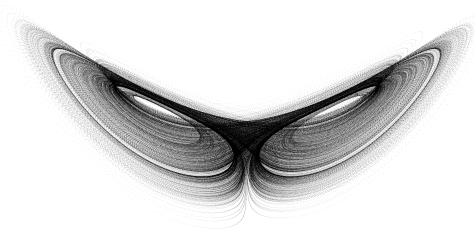


Figure 3. L'attrattore di Lorenz, proiettato sul piano  $zx$

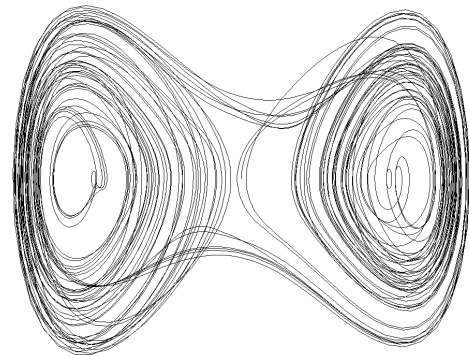


Figure 4. L'attrattore di Duffing,

sistemi dinamici non-lineari tra orbite periodiche e caotiche, in modo tale che anche il comportamento del sistema variasse tra il carattere periodico, quasi-periodico e caotico. In conclusione ai suoi studi, osservò che i brani generati a partire da sistemi con dimensioni maggiori, quindi con maggior numero di variabili attribuibili ai diversi parametri della data nota, sebbene risultassero maggiormente coesi a livello strutturale e formale, non manifestavano un risultato musicale di maggior interesse [16].

L'anno successivo, Agostino Di Scipio si cimentò in entrambi gli approcci, sia ad alto livello sia a basso livello, in due differenti lavori. In *Fractus*, brano per viola e suoni elettronici (1990), si focalizzò sulla generazione simbolica e sulla forma macrostrutturale della composizione. Invece di applicare i valori ottenuti dai sistemi caotici direttamente ai parametri di una singola nota, li utilizzò per concatenare gesti sonori predefiniti:

“un insieme numerato di elementi musica-

li caratterizzati come comportamenti timbrici particolari, cioè già considerabili come simboli nella articolazione dell'intera costruzione [14].”

Nel 1991, Michael Gogins lavorò invece sui sistemi di funzioni iterate [17], legati alla generazione di strutture frattali, che applicò nella composizione del brano acusmatico *Chaotic Squares*. Come egli stesso spiega, questi sistemi sembrano essere

“more powerful than previous methods for generating musical scores from fractals [18].”<sup>6</sup>

Richard Bidlack pubblicò l'anno seguente un articolo [15] sulla “semplicità e complessità” dei sistemi dinamici non-lineari. Egli compose una raccolta di canoni basati su diversi sistemi caotici multidimensionali descritti dalle equazioni di Hénon e di Lorenz, i cui valori venivano applicati a uno o più parametri di una stessa nota. Bidlack si concentrò su una particolare differenza tra questi sistemi: esattamente come i sistemi fisici, i sistemi dinamici possono essere dissipativi o conservativi. Egli notò come i sistemi dissipativi tendano ad assestarsi su un moto quasi-periodico, mentre quelli conservativi tendano a mantenere un comportamento più disordinato. La *sensibile dipendenza dalle condizioni iniziali*, ovvero la proprietà di questi sistemi di variare in modo considerevole il loro comportamento nonostante valori di partenza molto simili, ha un ruolo preponderante nella scrittura dei brani.

Negli anni successivi, l'applicazione dei sistemi caotici alla composizione musicale ebbe una diffusione sempre maggiore. Nel 1994, Mikhail Malt sviluppò *Chaos-librairie de modèles chaotiques et de fractales*, una libreria incentrata sullo sviluppo di sistemi caotici per il programma di composizione assistita Open Music. Tale libreria metteva a disposizione di tutti i compositori, anche i meno avvezzi alla Teoria del Caos, la possibilità di cimentarsi nell'utilizzo di equazioni caotiche. Alla fine degli anni Novanta l'uso di sistemi caotici e di frattali per generare musica diventò prassi diffusa e si estese anche ad altri generi musicali [19].

Nel 2004, André Sier pubblicò la prima libreria per il software Max/Msp chiamata A-Chaos Lib, in cui implementò più di una ventina di sistemi dinamici caotici caratterizzati da particolari attrattori [20].

È ancora degno di nota il lavoro di Jeremy Leach, che creò un programma di composizione automatica, chiamato XComposer, che utilizza la relazione frattalica  $1/f$  per generare materiale melodico. È stato statisticamente dimostrato che tale distribuzione probabilistica di generazione casuale di altezze risulta la più gradevole all'ascolto, poiché si muove su intervalli che sono i più comuni nelle melodie occidentali [11]. Tuttavia, poiché il *fractional noise* è un processo stocastico e comporta variabili casuali, non è possibile creare comportamenti regolari o ripetitivi, cosa che invece accade spesso a livello melodico. Leach basò quindi il suo software su equazioni caotiche,

<sup>6</sup> “più potenti dei metodi precedenti nella generazione di partiture musicali da frattali”, traduzione degli autori.

per poter generare ripetizioni di sequenze di altezze, ritmi e ampiezze [21].

Anche gli automi cellulari furono molto impiegati ad alto livello nella composizione automatica. Il fenomeno dell'*emergenza* è forse la caratteristica di maggior fascino di questi sistemi, nonché la più emblematica. Gli automi cellulari si sono rivelati una formidabile risorsa per i compositori: dal *Game of Life* al *Demon Cyclic Space*, questi sistemi hanno avuto molta fortuna nella rappresentazione di strutture sopraccellulari, generate dal fenomeno dell'*emergenza*, applicate alla generazione di materiale musicale [22].

### 3.2 Applicazioni a basso livello

L'approccio a basso livello di applicazioni degli algoritmi caotici alla produzione musicale si concentra invece sulla sintesi sonora e sul controllo del segnale audio.

Nel 1990 Barry Truax, in velata polemica con i lavori proposti fino a quel momento, sostenne che le applicazioni del caos nei processi compositivi ad alto livello generavano materiale musicalmente poco interessante:

“From a more philosophical or aesthetic point of view, it is not clear that an arbitrary mapping of a non-linear function [onto the pitch of successive notes] is inherently more musical than, for instance, a random or stochastic function. The musicality may reside in the musical knowledge of the mapper more than in the source function. The audience, if suitably primed with program notes, may be convinced there is more value or interest in the result because of the technique used, but the half-life of such interest seems to be short.”<sup>7</sup>

Tuttavia, ritenne che:

“a more inherently micro-level implementation seems desirable if flexible timbre generation is to be performed [12].”<sup>8</sup>

Anche Agostino Di Scipio, nello stesso anno di Truax e parallelamente alla composizione di *Fractus*, utilizzò sistemi dinamici non-lineari di Lorenz per controllare i parametri della sintesi granulare. In particolare, utilizzò le traiettorie tracciate dal sistema caotico sulla dimensione temporale del singolo grano: in questo modo ottenne suoni il cui spettro è in continua evoluzione, mai uguale a se stesso, senza perdere tuttavia una certa omogeneità e periodicità.

L'anno precedente, nel 1989, Rodeny Waschka e Alexandra Kurepa applicarono sistemi ricorsivi alla generazione di figure frattali al controllo di parametri della sintesi granulare o al controllo via MIDI di un sintetizzatore

<sup>7</sup> “Da un punto di vista più filosofico o estetico, non è chiaro se una mappatura arbitraria di una funzione non lineare [per esempio tradotta in una successione di note] è intrinsecamente più musicale rispetto, ad esempio, a una funzione casuale o stocastica. La musicalità può risiedere maggiormente nella conoscenza musicale del compositore che nella funzione di origine. Il pubblico, se opportunamente edotto con le note di programma, potrebbe essere convinto che ci sia più valore o interesse a causa della tecnica utilizzata, ma il tempo di dimezzamento di tale interesse sembra essere breve”, traduzione degli autori.

<sup>8</sup> “un'implementazione più inerente al microlivello [del suono] sembra essere preferibile qualora si ricerchi una generazione di timbri flessibile”, traduzione degli autori.

Synclavier II. Uno dei problemi che riscontrarono nella so-  
nificazione dei frattali è che, a differenza di quanto accade  
a livello percettivo visivo, l'ascoltatore non può avere la  
visione completa della struttura, come accade guardando  
una figura frattale nel suo insieme [23].

Ancora, nel 1995, John Mackenzie studiò vari algoritmi  
di tipo caotico applicati alla sintesi, basandosi sul concetto  
di feedback e autocontrollo. A partire da uno studio sull'analisi  
e risintesi dei suoni, notò come gli algoritmi caotici,  
se utilizzati per la generazione di involucri nella dinamica  
dei suoni, rendessero il suono sintetizzato più vivo e naturale  
di quanto potessero fare altri algoritmi deterministici.  
I sistemi dinamici non-lineari, infatti, in molti casi possono  
rispecchiare un modello fisico o derivare da esso [24].

### 3.3 Oscillatori caotici e l'oscillatore di Chua

Un altro approccio che si può considerare a basso livello,  
sebbene sia sostanzialmente differente dalle applicazioni  
sopracitate, è lo sviluppo di *oscillatori caotici*. A tal ri-  
guardo, occorre citare un altro sistema caotico molto im-  
portante e modello di riferimento per lo sviluppo di questi  
oscillatori, ovvero *l'oscillatore di Chua* [25].

L'oscillatore di Chua è basato su un circuito analogico  
molto semplice (si veda Fig. 5), la cui caoticità è deter-  
minata da un particolare diodo, detto *diodo di Chua*, che  
manifesta caratteristiche non-lineari, non solo in funzione  
della frequenza di oscillazione, ma anche in relazione alle  
altre componenti (resistenze e condensatori) presenti nel  
circuito. Al variare dei parametri relativi a queste compo-  
nenti, il circuito presenta un insieme di soluzioni in cui si  
alternano regioni di periodicità (oscillazioni regolari) e re-  
gioni di instabilità (oscillazioni caotiche), convergendo su  
un punto fisso, oppure su un ciclo limite o ancora su un  
*strange attractor* [26].

Le equazioni che descrivono l'oscillatore di Chua sono:

$$\begin{cases} \dot{x} = p_1(ky - x - f(x)) \\ \dot{y} = p_2(kx - y + z) \\ \dot{z} = -q(y + rz) \end{cases} \quad (1)$$

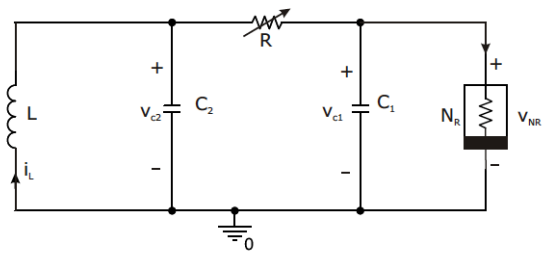
dove,

$$f(x) = \begin{cases} bx + (a - b) & x \geq 1 \\ ax & |x| < 1 \\ bx - (a - b) & x \leq -1 \end{cases} \quad (2)$$

con  $p_1, p_2, q, r$  e  $k$  parametri relativi alle componenti e  
caratteristiche del circuito e  $a, b$  parametri relativi al diodo  
di Chua.

Focalizzando l'attenzione sulle variazioni qualitative del-  
le dinamiche dell'oscillatore, al variare del solo parametro  
 $r$  che costituisce il parametro di biforcazione, una volta de-  
finiti i valori degli altri parametri, si osservano diversi modi  
in cui l'oscillatore passa da un comportamento regolare a  
uno caotico:

- per il *raddoppiamento del periodo*, ovvero il sistema  
si assesta su un ciclo limite il cui periodo tende a rad-  
doppiare all'aumentare del parametro  $r$ , diventando  
sempre più articolato, fino a raggiungere la soglia



**Figure 5.** Il circuito raffigurato rappresenta uno dei  
modelli più semplici del circuito di Chua.

critica, oltre la quale non è più possibile definire un  
periodo;

- per *intermittenza*, ovvero per determinati valori di  $r$   
il segnale manifesta ad intermittenza comportamenti  
periodici e comportamenti aperiodici;
- per *biforcazioni per aggiunta di periodo*, ovvero si  
genera un secondo ciclo quasi-periodico che si alterna  
in maniera caotica con il primo ciclo.

Sono stati molti i tentativi utilizzare questo circuito alla  
sintesi sonora e ad un utilizzo musicale [27]. Nonostante  
la grande attrattiva e curiosità a riguardo, emersero proble-  
matiche relative alla stabilità e al controllo di questo stru-  
mento, limitandone l'applicazione pratica.

Altri esempi di oscillatori caotici - di natura principal-  
mente digitale - consistono nella modellizzazione di siste-  
mi dinamici non-lineari o altre forme di sistemi ricorsivi  
(tra i quali anche i già citati Lorenz o Duffing) e applicarli  
alla generazione di segnale audio, campione per campione.  
La peculiarità di questi oscillatori sta nel loro comporta-  
mento: con estrema economia di mezzi è possibile ottenere  
oscillazioni regolari e periodiche, generando quindi suoni  
armonici con altezze definite e spettri relativamente sem-  
plici, e oscillazioni irregolari e caotiche, generando invece  
suoni inarmonici, spettri molto più complessi. Il compor-  
tamento di tali sistemi è infatti analogo all'oscillatore di  
Chua, almeno nella modellizzazione informatica.

Shahrokh Yadegari, nel 2003, pubblicò un interes-  
sante studio [28] sugli oscillatori caotici, derivati dal modello  
di Lorenz e dal modello di Duffing, affiancati all'im-  
plementazione dell'oscillatore di Chua. Gli oscillatori caotici  
suscitarono fin da subito grande interesse per le loro poten-  
zialità e la loro versatilità, tuttavia il loro impiego a livello  
compositivo è stato relativamente scarso.

## 4. CONCLUSIONI

Come abbiamo visto, sono state molte le applicazioni del-  
la Teoria del Caos nella *computer music*, come copiosa la  
ricerca e il dibattito intorno all'argomento. La concezione  
del caos ha avuto negli ultimi vent'anni ulteriori e interes-  
santi sviluppi. Grazie alla crescita esponenziale della po-  
tenza di calcolo dei computer, ci si è spinti sempre più a  
fondo nella modellizzazione dei sistemi caotici.

Sebbene con il recente interesse per gli algoritmi di *machine learning* il tema del caos deterministico e le implicazioni teoriche che comporta siano passati in secondo piano, negli ultimi anni sono stati molti i compositori che si sono avvicinati al caos deterministico e molti gli studi e le pubblicazioni sull'argomento.

Ancora oggi, infatti, il caos deterministico ispira compositori a confrontarsi con le recenti tecnologie: ne è esempio John Mackenzie, che nel 2013 ha creato un sintetizzatore basato su equazioni caotiche distribuito sotto forma di applicazione per sistema iOS chiamata "Strange Attractors" [24]; o Miroslav Spasov, che ha pubblicato nel 2015 una libreria per il programma Max/Msp chiamata anch'essa "Strange Attractors" [29], in cui implementa una decina di algoritmi finalizzati all'utilizzo in tempo reale.

Inoltre, algoritmi di natura caotica mantengono tuttora un vivo interesse in ambiti musicali paralleli a quello più strettamente compositivo come, ad esempio, in pratiche performative, interattive e improvvisative [30].

## 5. REFERENCES

- [1] J. Gleick, *Caos, La nascita di una nuova scienza*. Viking Penguin Inc., 1987.
- [2] P. de Laplace, *Théorie analytique des probabilités*. Courcier, 1814.
- [3] F. Berto, *Tutti pazzi per Gödel!: la guida completa al teorema di incompletezza*. Laterza, 2009.
- [4] D. Ruelle, *Caso e Caos*. Editions Odile Jacob, 1991.
- [5] S. Wiggins, *Introduction to Applied Nonlinear Dynamical Systems and Chaos*. Texts in Applied Mathematics, Springer New York, 2006.
- [6] J. Mishra and S. Mishra, *L-System Fractals*. Mathematics in Science and Engineering, Elsevier Science, 2007.
- [7] S. Wolfram, *Cellular Automata And Complexity: Collected Papers*. CRC Press, 2018.
- [8] L. Morino, "Caos deterministico: la teoria del caos nella computer music." Tesi di Diploma di primo livello presso il Conservatorio "G. Verdi" di Torino, 2016.
- [9] C. Roads, *Computer Music Tutorial*. MIT Press, 1996.
- [10] C. Dodge and T. A. Jerse, *COMPUTER MUSIC Synthesis, Composition and Performance*. Springer-Verlag/Wien, 1997.
- [11] R. Voss and J. Clark, "1/f noise in music: Music from 1/f noise," *Journal of the Acoustical Society of America*, 1978.
- [12] B. Truax, "Chaotic non-linear systems and digital synthesis: An exploratory study," *Proceedings of the 199 International Computer Music Conference*, pp. 100–103, 1990.
- [13] C. Dodge, "Profile: A musical fractal," *Computer Music Journal*, vol. 12:1, pp. 10–14, Spring 1988.
- [14] A. Di Scipio, "Caos deterministico, composizione e sintesi del suono," *Atti del IX Colloquio di Informatica Musicale*, pp. 337–350, 1991.
- [15] R. Bidlack, "Chaotic systems as simple (but complex) compositional algorithms," *Computer Music Journal*, vol. 16:3, Fall 1992.
- [16] J. Pressing, "Nonlinear maps as generators of musical design," *Computer Music Journal*, vol. 12:3, pp. 35–46, Fall 1988.
- [17] M. Barnsley and S. Demko, *Iterated Function Systems and the Global Construction of Fractals: Proceedings of the Royal Society of London*. Proceedings of the Royal Society of London: A, Royal Society, 1985.
- [18] M. Gogins, "Iterated functions systems music," *Computer Music Journal*, vol. 15:1, pp. 40–48, Spring 1995.
- [19] M. Malt, "Lambda3.99 (chaos et composition musicale)," *Troisimes Journes d'Informatique Musicale (JIM 96)*, 1996.
- [20] A. Sier, "A-chaos lib." <http://s373.net/code/A-Chaos-Lib/A-Chaos.html>, 2004.
- [21] J. Leach and J. Fitch, "Nature, music, and algorithmic composition," *Computer Music Journal*, vol. 19:2, pp. 23–33, Summer 1995.
- [22] K. McAlpine, E. Miranda, and S. Hoggart, "Making music with algorithms: A case-study systems," *Computer Music Journal*, vol. 23:2, pp. 19–30, Summer 1999.
- [23] R. Waschka and A. Kurepa, "Chaotic non-linear systems and digital synthesis: An exploratory study," *Proceedings of the 1989 International Computer Music Conference*, pp. 332–335, 1989.
- [24] J. P. Mackenzie, "Chaotic predictive modelling of sound," 1995.
- [25] X. Rodet, "Sound and music from chua's circuit," *Journal of Circuits, Systems and Computers*, pp. 49–61, 1993.
- [26] L. Pivka, C. W. Wu, and A. Huang, "Chua's oscillator: A compendium of chaotic phenomena," *Journal of the Franklin Institute*, 1995.
- [27] G. Mayer-Kress, I. Choi, N. Weber, B. R., and H. A., "Musical signals from chua's circuit," 1993.
- [28] S. Yadegari, "Chaotic signal synthesis with real-time control: solving differential equations in pd, max and jmax," in *Proceedings of the 6th International Conference on Digital Audio Effects*, September 2003.

- [29] M. Spasov, “Using strange attractors to control sound processing in live electroacoustic composition,” *Computer Music Journal*, vol. 39:3, pp. 25–45, Fall 2015.
- [30] T. Mudd, *Nonlinear dynamics in musical interactions*. PhD thesis, Music Computing Lab, Centre for Research in Computing, The Open University, 2017.

# To embody or not to embody: A sound design dilemma

**Stefano Delle Monache**

Iuav University of Venice  
 Dept. of Architecture and Arts  
 stefano.dellemonache@iuav.it

**Davide Rocchesso**

University of Palermo  
 Dept. of Mathematics and Computer Science  
 davide.rocchesso@unipa.it

## ABSTRACT

Designing with sound is about constructing appropriate sound representations of a concept, from the early ideas to the final product. A survey research on embodied sound sketching is presented, and the problems of early representation in sound design are discussed by analysing the questionnaire results of three workshops on vocal sketching.

## 1. INTRODUCTION

Is there a difference between sketching sound and sketching with sound? What is a sound designer expected to do when approaching from scratch a creative project involving sound? It has been argued that sound design is a funnel-shaped iterative process of analysis, creation and evaluation [1], and that the sound designer operates at the intersection of several disciplines (acoustics, psychoacoustics, music studies, psychology and computer science). This professional is an artist/composer who has the *tacit knowledge* to mediate the early industrial, perceptual and physical constraints into aesthetic sound specifications [2]. On the contrary, according to [3], the sound designer is essentially an engineer provided with a strong and interdisciplinary skillset, able to carry out the product sound development task, autonomously. Somewhat in the middle, Erkut and colleagues situate sound design in the wider context of multisensory product appraisal, and look at interaction design and product orientation as reference domains [4]. As such, sound represents one opportunity to alter the perceived product attributes, through design choices. Along the same line, Hug and colleague advocate a sound-driven interaction design approach by means of performance-led, increasingly complex and refined representations [5].

However, at the beginning of any project, the designer has to confront with a given brief, analyse and set the problems, conceptualise abstract (i.e., values, beliefs, principles) and concrete (i.e., sensory, physical, spatial, temporal) attributes, and progressively embody them into representations, from early ideas and mock-ups, to prototypes

up to the final product [6]. Would it be more appropriate to talk about designing (also) with sound, then? We believe that there is a substantial shift of perspective, not only because the aural dimension is taken into account from scratch, together with the other aspects (visual, tactual, formal, etc.), but also because audition emerges as one of the channels through which materialising conceptual associations throughout the design process. The consequent question is which kind of representation is more effective, at which stage of the process, to embody concepts in and through sound.

Designing *with* sound puts *design thinking* at the center of the sound creation activity, whereas designing sound seems to suggest rather a focus on sound production of quality factors in the refinement and optimisation of a product. In other words, designing sounds rather reflects the sonic representation *per se* in its formal elements, assimilable to a *reduced listening* approach. According to Visser's proposition, we see design thinking in the aural domain as a peculiar implementation of a generic cognitive activity (i.e., design as construction of representations), and yet clearly distinct from other cognitive activities (e.g., economics) [7].

In this paper, we problematise and reflect on this apparently controversial aspect, by reporting the qualitative assessment of three workshops on sonic sketching. Therefore, we consider only the early stage of the creative process, wherein ideas generation and conceptualisation take a fundamental and considerable part. What we strive to understand is the focus of sound design as a reflective practice, from an embodied cognition perspective, that is considering the body as the natural mediator between internal, mental representations of concepts and external representations in sounds [8]. As such, carving the auditory information and sensation (i.e., designing sound) is only one type or step of product embodiment, among e.g., in a non-exhaustive list, anthropomorphism, symbolic meaning attribution, familiarity, agency and action-sound coupling [9, 3, 10].

We contend that the problem of sound design is essentially a problem of construction of representations, "in an area where best practice is virtually inexistent" [5]. We mean, for instance, that types of visual representations are existing and taxonomically classified, depending on the design stages, the design or technical information, the function and the detailing of the representation (e.g., a study sketch, a 3D model, or a prototype), and are widespread in the professional practice [11]. The same depth of knowl-

Copyright: ©2018 Stefano Delle Monache et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution License 3.0 Unported](https://creativecommons.org/licenses/by/3.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

edge and expertise is not available yet in the realm of sonic representations of designs [6].

Section 2 summarises the basic characteristics of conceptual sound design within the framework of embodied cognition, with a special focus on sketching through vocalisations; Section 3 introduces the workshop format on vocal sketching, and the evaluation questionnaire; comparative analysis and results of three workshops are reported and discussed in Section 4.

## 2. EMBODIED SOUND SKETCHING

Sketching as a product and a process is that peculiar design activity through which designers externalise their ideas into representations. Typically, the word sketch is associated to drawing, which is the first and most immediate spatial way to embody concepts held in mind, into a visual artefact. This is only the first leg of a journey in the design funnel, which leads the initial concept(s), existing in one's own body and mind only, to being distilled and embodied into structural properties and configurations of the design entity under scrutiny.

Drawing is an effective means because it exploits a tool, the pencil, and a spatial support, the paper, that allow a speed of production coherent with the immediacy and flow of thought. It is a form of expression tolerant to incompleteness and inaccuracy, it is reversible, and yet capable to provide evocative, economical and disposable impressions. The body, by means of the hand, is actively engaged in sketch-thinking by encoding and exploring the information [12]. The body at large, through talk, action, and gesturing, is all the more so involved in catching and expressing the dynamics of situations, processes, systems, and behaviours, that do not exist in space only, but especially in time [13, 14, 15].

Hence, sketching does not equal to drawing, it is rather a peculiar way of thinking and searching which is independent from the medium, e.g., a drawing, a photograph, a pantomime, physical props, and sound. What matters is the appropriateness of the chosen medium in representing, evoking and possibly sharing a specific concept. Figure 1 represents sonic representation techniques and methods in the designing-with-sound funnel, according to their degree of embodiment and detailing [17].

A primary viable way to sketch concepts is to browse and wander through collected representations of direct percepts. Designers are prone to collect, annotate, and archive external images in sketchbooks, for a variety of purposes. Archives are usually organised according to subjective categorisations, which take in account distal (source-related), proximal (sensory-related) or semantic representation strategies. In the early stage of the project, sound designers largely resort to "external" images and visual analogies to illustrate the overall concept of their idea [18, p.35]. Sounds collections are mostly kept in the form of sound banks and libraries, which respond to the rationale of an efficient selection task in the workflow of Digital Audio Workstations (DAW), rather than a serendipitous encounter and exploration of others' works.

In our research through design workshop activities, we

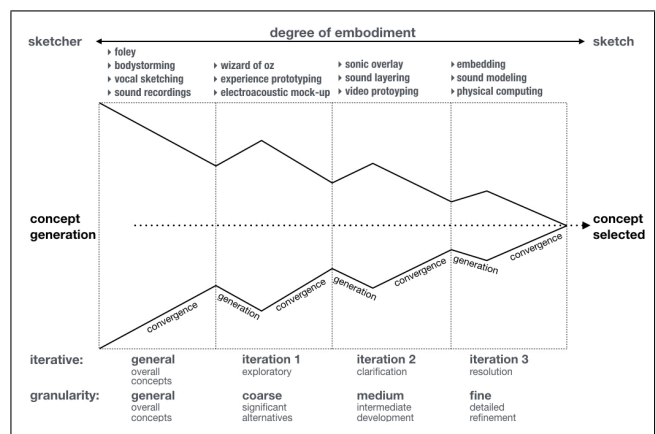
have been striving to subvert the prevalent disembodied sound design practice and recover the original rationale of personal collections, namely having thoughtful materials that facilitate ideas generation and reflection, rather than sound materials to be readily edited and processed [19]. As an example, two introductory "sound hunting" exercises may play as follows:

**Exercise 1:** Go out and hunt sounding objects in the wild. Collect them (physically!) and write down the reasons why sound captures your attention. Revise the text and try to label the reasons according to sound categories and conceptual associations. Skim the most meaningful label(s) representing the sounding object and start populating the corresponding boxes.

**Exercise 2:** Go hunting a sound event and make the audio-visual recording. Try to understand the reason(s) why that particular sound idea interests you. Go out in the wild again, and look for 20 different samples that follow your idea.

Now, work by contradiction, and look for sound events that evoke the opposite of your idea, e.g., dynamic - static, hot - cold, happy - sad, energised - exhausted, textural - gestural, informative - confusing, etc.

Here, Foley-oriented practices [5, 20] are aimed at fostering the construction of personal thoughtful collections. A second way to sketch concepts is to actually externalise ideas and "images" held in the mind, and thus producing self-generated representations. These include *performed* sounds, that is made with an *intention*, whether through vocalisations or physical manipulation of objects. In embodied sound design, access to imagery is rooted in the sensory-motor nature of auditory experiences. Through his body, the sound designer accesses such internal representations, previously stored as perception - action ensembles [8, 10, 21].



**Figure 1:** Generation, elaboration and reduction of sound ideas represented as a funnel, with diverse methods suitable to "visualise" the sonic concepts according to the degree of embodiment centred on the sketcher or the sketch (adaptation from [16, p.11]).

## 2.1 Vocalisations as embodied sketches

The vocal apparatus is the embodied sound producing tool par excellence, available to sound designers. Recent research on perception, articulation and production of sounds shed light on the inner mechanisms involved in the communication of concepts via vocal and gestural imitations [22, 23, 24]. Vocal sketching is the embodied design method to represent concepts, by means of vocalisations [25]. More recent design research on vocal sketching has been systematising this form of expression in well-defined exercises and practices [19].

Vocal imitations are the closest semantic representations, directly ascribable to a person's idea of a given sound. They are more effective than verbal descriptions of sounds, especially when speakers run out of words, and rely on subtle sensory processing mechanisms, that is the selection of the most relevant acoustic features, for the identification of a given sound, that can be produced with the human voice. It has been shown that the human voice can reproduce the pitch, the rhythm, the sharpness (i.e., spectral centroid) of a given sound [24]. Vice versa, when vocal imitations are elicited from memory (i.e., imagery), they rely on some sort of iconic similarity with the thought sound, that is a few salient characteristics picked up and shifted in the imitator's vocal register [23]. In this respect, vocal imitations act much like sound caricatures. Vocal imitations are sketches of auditory concepts, in the same way freehand drawings represent visual objects.

Finally, computing systems aimed at supporting such immediacy of communication and reflection through sound creation, by returning synthetic representations (i.e., configurations of digital sound models) as instances of vocalisations, are making progress. In [21], we showed how SEED, an embodied-sound-design tool, affords to produce, with a certain degree of reliability, tamed and predictable representations, grounded in vocal motor skills and control: sound designers were able to reproduce, with SEED, target examples previously created by a third sound artist-designer, with the same tool. A relevant aspect of embodied sound representations is that they are naturally prone to coordination and conflict resolution across internal mental models, and thus available to negotiation and collaborative practices, similarly to collaborative drawing on a shared whiteboard [26].

## 3. DESIGN RESEARCH THROUGH VOCAL SKETCHING

To our knowledge, the first design research workshop on the methodology of vocal sketching was organised in 2009 by Ekman and Rinott, as part of the COST Action on Sonic Interaction Design [27]. The outcome of that workshop represents a seminal study in the area of sonic interaction design [25]. The two authors conceived a set of warm-up and design tasks in collaborative vocal sketching of sonic interactions, that were assessed through ethnographic observations and post-workshop questionnaire with open-ended questions. In particular, the study aimed at evaluating the usefulness of sketching with voice, and whether

and in which way the proposed method may influence the design process. The main findings of the qualitative assessment of the design tasks and process can be summarised as follows:

1. the vocal apparatus provides “instant access to sonic reserves [of complex, organic sounds] that are hard to achieve with current tools, and away from simplistic sound solutions”, such as beeps and bleeps [25, p.128];
2. voice is by no means a neutral tool, and yet the inherent limitation and constraints may influence decision-making in design directions, positively and negatively;
3. being the main limitations, the monophonic and harmonic character of the voice, the difficulty in producing and controlling specific, complex sounds, the sound duration due to limited breath cycle;
4. on the other side, vocal sketching as a method does not affect the design process, and provides a rich variety of sonic expressions;
5. effective vocal sketching requires practice and fluency.

Since 2009, vocal sketching as method to support conceptual sound design has been explored in a variety of contexts, was included as reference resource in sonic interaction design education and research, and fostered the development of several voice-driven interactive technologies [28, 29].

### 3.1 Three workshops on vocal sketching

In the 3-year period (2014 - 2016), we conducted extensive design research on vocal sketching, within the scope of the EU project SkAT-VG (Sketching Audio Technologies using Vocalisations and Gestures) [30]. The workshop was the preferred venue in which *i*) exploring and understanding use strategies of voice and gestures for design purposes, *ii*) studying the sound designers' behaviour in the creative setting, *iii*) developing a propaedeutics on embodied sound sketching, through exercises grounded in phonetics (i.e., elicitation and articulation), and auditory perception of vocal imitations, *iv*) proposing and assessing voice-driven tools for embodied sound design [31, 21].

A workshop format on vocal and gestural sketching emerged from the organisation of several workshops. We refer to the extensive report of the research through design activities and exercises, provided in [18, p.10]. The workshop is normally split in two main parts: A first preparatory phase aims at sensitising the participants to sonic interaction and warming up the use of the voice, and a second phase rather focuses on proper sound sketching and design tasks. Shared doing, reflective practices and inter-observation are instrumental to enable understanding through designing sonic and interactive objects, and 10 – 20% of frontal teaching is balanced with 90 – 80% of hands-on, learning by doing activities. The resulting framework can be scaled from one-day to one-week duration.

After each SkAT-VG workshop, a questionnaire was handed out to the participants, in order to collect feedback

on their experience with sound and vocal sketching, and on the effectiveness of the support software tools under development.

In this paper, we present and discuss the results emerging from the analysis of the questionnaires collected in three workshops. Two workshops took place in November 2015: The first (W1) was a two-day edition organised at the University of York, Department of Theatre, Film and Television, with seven participants. The second workshop (W2) was a five-day edition organised at the Aalborg University of Copenhagen, in the scope of the Master in Sound and Music Computing, with twenty participants. The third two-day workshop (W3) took place in March 2018, at the Conservatory of Music “G. Verdi” of Torino, in the scope of the Master in Electronic Music and Technologies of Sound, with eleven participants.

The general background of the participants, across the three workshops, was quite homogeneous, ranging from music composition and sound production for linear media (sound effects, Foley, soundtrack), to sound design for videogames and interactive installations. Yet, no participant had prior experience or formal education in design and product sounds. W1 was essentially based on vocal sketching practices, with no support of voice-driven software tools. In W2, we introduced a sound sketching session using “miMic”<sup>1</sup>, a voice-driven environment in which physics-based sound models [32] can be selected as instances of previously classified vocal imitations, and further controlled through vocalisations [31]. miMic provided the skeleton for the development of SEeD, a more advanced system for sketching synthetic sound representations, based on physics-based sound models and corpus-based synthesis [21]. SEeD<sup>2</sup> was instead introduced in W3.

### 3.2 The questionnaire

The SkAT-VG questionnaire<sup>3</sup>, based on the one conceived by Ekman and Rinott in [25, p.131], is arranged in two main parts. The first part contains close-ended, 7-point scale questions on the design effectiveness and appraisal of vocal sketching, and on the general relevance of other forms of sound sketching including the physical manipulation of objects and embodied sound synthesis. From the latter viewpoint, the questions are specifically addressed at assessing the immediacy of use and expression of the Sound Design Toolkit<sup>4</sup> (SDT), a set of physically-informed, procedural sound models for the Max environment, developed by SkAT-VG [32], and SEeD tool, in which configurations of SDT models are driven by vocal input and control [21]. The second part, in the form of open-ended questions, encompasses general comments on previous experiences in vocal sketching, and its usefulness and gaps in comparison with Foley and SEeD.

<sup>1</sup> miMic demo: <https://vimeo.com/142351022>.

<sup>2</sup> SEeD in action: <https://vimeo.com/271826511>; <https://vimeo.com/271825753> (0:50).

<sup>3</sup> Questionnaire available for download at <https://owncloud.skativg.iuav.it/index.php/s/0rB11NS4QyCvz31/download>.

<sup>4</sup> <http://soundobject.org/SDT>.

In the following section, we reflect on the picture emerging from the analysis of the answers to the range questions, and the written comments collected. The analysis of videos and design outcomes per workshop is beyond the scope of this paper.

## 4. RESULTS AND DISCUSSION

The results of the survey are represented in Figure 2. The bar charts report the answers relative to the three main topics of the questionnaire. The median ratings by participants per workshop are arranged in a row, and in a chronological order from left to right, to ease both the reading of the emerging picture, and the comparison between the three workshops. Note that in Figure 2c, the W1 chart does not include the last three questions (Q12 – 14), as software support to vocal sketching (SEeD) was not available yet.

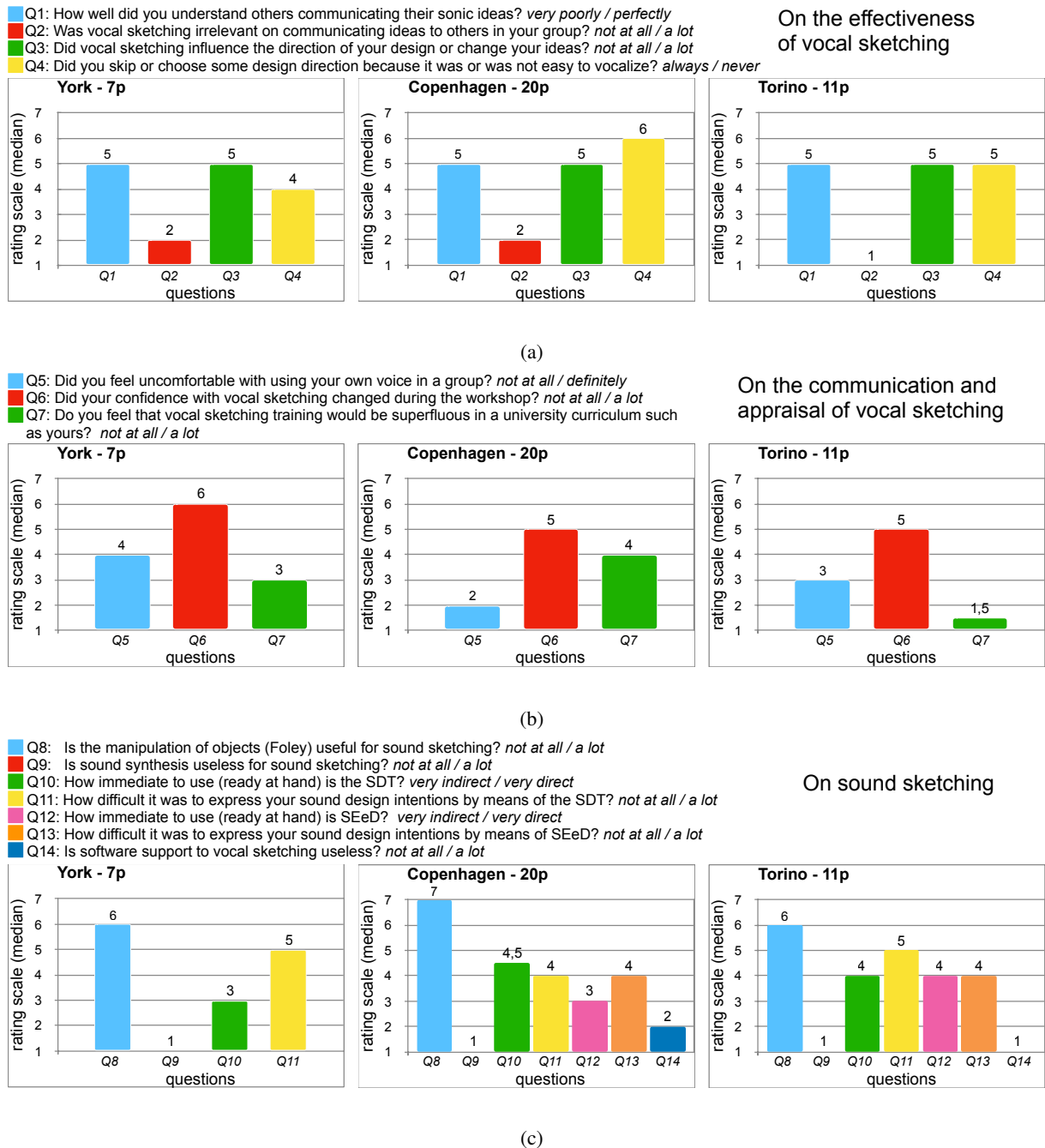
### 4.1 On the effectiveness of vocal sketching

From the inspection of Figure 2a, it emerges a relatively good commitment of participants to the exploration of voice as means to communicate sound ideas (Q1/Q2). Representing imagined sound by means of individual or collaborative vocalisations is generally found effective. This is corroborated in the open answers, according to which sketching with voice was found useful for a variety of reasons: It emerges that, based on the participants background, the proposition of 1) considering sketching in sound creation, 2) considering an embodied approach to early representations, 3) using voice as sketching tool, was reflected in diverse understanding of opportunities.

**Influence on the design process:** Participants with prior experience in interactive, cross-media installations reported that sketching with voice is “interesting and helpful for interaction design” and “product sound”, it is “fast and efficient, and not bound to language barrier”, it “can help to think out of the box”, and “provides an intuitive and explicit way to adapt to experimentations and obtain simulations”, “free from hardware and software constraints”. The interest is focused on the immediacy and speed of production, intrinsic to vocalisation, in designs characterised by inherent perceptual relations in temporal dynamics.

**Vocalisations as communication tool:** Participants with major prior experience in sound effects for linear media and video games highlighted the sharing aspects of early drafts made out of vocalisations, when ideas need to be communicated to peers, especially in teamwork. They found vocal sketching useful “to express my sound design ideas in a more direct and clear way”, “to describe the sound in another way”, “to convey ideas on how something should look like”, “when we wanted to make some sound, but we couldn’t find one appropriate”. In particular, vocal sketches, both as process and product, are seen as means “to quickly produce video prototypes in a group”, and especially “useful in the pre-production stage”. However, several participants stressed that the great effort required to articulate continuous sounds, due to the limited breath capacity, severely affected the sense of continuity.

**Voice as sound generator:** This is the most apparent, and



**Figure 2:** Bar charts with the median of the answers to the range questions, arranged per topic, and workshop.

misleading understanding of sketching with voice, which reflects the established cast of mind in sound creation and design. Vocal sketching “is an extra tool to use”, suitable to produce “brief, immediate and general effects”, such as “impact sounds”, “noises” and pitched mechanisms (e.g., engines), and “to create atmospheres”. On the other side, common comments on the deficiencies of voice are the difficulty in “creating futuristic and machine-like sounds”, producing “high-pitched, morphologically complex” (e.g., “metallic and large”, “low and dense”), “refined and outer” sounds that “do not have an equivalent in the real world”, “obtaining smooth changes in sounds”. In addition, “vocal sketches sound often too similar to onomatopoeia”, and “voice drew us towards the same kind of

sound design, drones, humming etc.”.

There are several considerations entangled in these answers, that we may group from three main arguments. The first is clearly rooted in the perception and cognition of vocal imitations, and auditory experiences in general, as discussed in Section 2.1. The second reflects the general expectation of having an all-in-one solution to produce sound, a sort of tool capable to interpret internal image of sound kept in the mind, and externalise it in its most complete form. Sound creativity in the digital age is much retained in the mind, by relying on mental synthesis processes, rather than enabling discovery in the observation-imagination continuum [12]. The dependence on software tools (e.g., DAWs, plugins, and programming environ-

ments) to create sounds can be ascribed to the idiosyncrasy emerging from the inherent need to fix a time-frequency representation on a meaningful support, and the creative, abstract (we prefer to say disembodied) approach to sound, inherited from the early computer music legacy [33].

Voice is certainly a sound generator, but the purpose of using it is not to model or implement a sound, in the same way the purpose of a drawing is not to physically build an object. Yet, it has to be noted that there are designs in which the distance between intermediary representations and the final product may vary dramatically. The design of an artefact is a different activity from its realisation [7]. In this respect, if the vocalisation is the closest semantic representation of a given idea of a sound, then the problem is that of the effective control of the sound generator, i.e., the vocal apparatus. An embodied approach to design requires fluency and expertise, but not virtuosity in the production of external representations. These are not mere reflections of images in the mind, they are proper thinking tools. Sketching and mental imagery go together, and the sound sketch represents a volatile, yet stable enough, external memory.

Answers to Q3 and Q4 much reflect this state of things. Vocal sketching as a method allows to progress in the design process, by means of non-verbal sound representations, despite the constraints and effort required to vocalise them. In this respect, these results are in line with the conclusions by Ekman and Rinott, according to which voice is by no means a neutral tool. Vocal sketches are effective externalisations of ideas and represent the first physical encounter with the image held in the mind, they are idea or thumbnail sketches meant for personal use [11].

As a final remark, participants with a background focused on music composition did not find vocal sketching very useful for their type of work. As prior experience in the use of voice, they reported the imitation of music parts, especially in “a cappella” music. This is plausible, since music sketching make large use of spatial representations (i.e., music notation, even abstract), often integrated by the direct support of the musical instrument. On the other side, the value of the musical sketch in the compositional process has been traditionally considered problematic, in the light of “the utopia of unhindered musical imagination developing on its own terms” [34, p.5].

Therefore, it comes at no surprise that the current generation of sound designers, essentially coming from music studies [2], are not acquainted with sketch-thinking [18, p.34]. In other words, despite the word “design”, a proper attitude and knowledge is still missing in the creation of aesthetic functional sound [26]. This motivates academic research in developing meaningful curricula for future generations [4].

#### 4.2 On the communication and appraisal of vocal sketching

The picture emerging from the answers in Figure 2b is that bodily involvement in communicating a design idea can be distressing. Nonetheless, the social discomfort in teamwork can be overcome through practice and shared activ-

ities (Q5/Q6). The hosting context of the workshops is reflected in the answers to Q7, about the convenience of a training in sketching with voice, within the curriculum of study. Considering the film sound production orientation of W1 participants, the rather engineering and technological attitude of W2 participants, and the artistic and music-oriented participation in W3, the usefulness of a training in vocal sketching received neutral or low scores, thus meaning that a training is not superfluous at all.

To test the hypothesis that the different scores in Q7 might reflect diverse appraisals of sketching with voice, a Kruskal-Wallis non-parametric test was run ( $H = 10.53$ ,  $p < .05$ ). Hence, there is actually a difference between the three groups, with respect to the attitude towards the use of voice as sound representation tool. We hypothesise that W1 participants may have mainly focused on the communication aspects, since no voice-driven tool was provided; W2 participants may have been attracted by software support opportunities; whereas W3 participants were rather involved in the creation process. The post-hoc pairwise comparison test revealed statistical significance only between W2 and W3.

Since the answers to questions 10 – 13 (Fig. 2c), on the expression and immediacy of use of the SDT and miMIC / SEeD, are essentially overlapping, we further investigated Q14 on the usefulness of software support to vocal sketching. We could not find any statistical significance in the difference between W2 and W3 (Mann-Whitney U,  $U = 56$ ,  $p = 0.0735$ ). Certainly, the minor interest by W2, in sketching with voice, cannot be ascribed to the fact that miMic<sup>5</sup> was apparently less ready-at-hand than SEeD (Q12, Mann-Whitney U,  $U = 40.5$ ,  $p = 0.49$ ).

A possible reason can be found in the answers to Q2, on the relevance of vocalisations as conceptual sketches. The Kruskal-Wallis test showed only a marginal difference between W1, W2, and W3 ( $H = 5.909$ ,  $p = 0.0521$ ). The post-hoc test reported a significant difference between W1 and W3 only. The pairwise Mann-Whitney testing of the three groups reported the statistical significance between W1 and W3 ( $U = 15.5$ ,  $p < .05$ ), and between W2 and W3 ( $U = 54$ ,  $p < .05$ ), but not between W1 and W2 ( $U = 51$ ,  $p > .05$ ). In other words, the different appraisal of vocal sketching, by W3, can be explained in terms of a higher involvement in the use of vocalisations to embody and communicate *concepts*, compared to the other two groups. W1 and W2 either resorted on other means to communicate ideas, that is verbal descriptions, sound samples, Foley, others, or focused on the use of voice as a mere sound generator. Hence, we can infer that W1 and W2 attitude to skip or choose a design direction, based on the complexity of vocalisations, rather reflects a similar behaviour, namely a lesser engagement in conceptualising the vocal representations. In our previous work [26], we showed how protocol and linkographic analysis of cooperative sound design sessions can provide relevant qualitative and quantitative information about the use of vocalisations in the creative process. A further consideration is

<sup>5</sup> miMic can be considered an early version of SEeD. In the practice, W2 and W3 made use of two different releases of the same software.

that while the exercises of the SkAT-VG workshop were still in an explorative fashion at the time of W1 and W2, W3 workshop could benefit instead of an overall refined program and pace.

### 4.3 On sound sketching

The bar charts in Figure 2c report the attitude of participants towards other complementary means to sketch sounds. As expected, the two main alternatives, that is physical manipulation of objects (Foley) and sound synthesis, scored a high preference (Q8/Q9). The open answers on the use of Foleys mostly focus on the expression and convenience of this sound production technique. A Foley approach is considered useful to make “sounds that are hard to access or obtain”, and especially “simple effects and unrelated sounds, not existing in the real world”, “for the interaction with the objects”. Foley, as a form of representation, was found effective when “creating quick prototypes” in “later stages of the sound design process”, because “hands and feet afford a more subtle control of the dynamics”. On the other side Foley sound production “takes too long” and “requires more teaching in technical tools”, since “it is difficult to produce a definite sound, kept in mind, if all the material properties are not well known and mastered”. In particular it is found difficult to create “subtle and ambient sounds” with “dynamic, highly variable events, which would be easier to synthesise.

Answers (Q10 – 13) on the immediacy of use and expression of physically-informed approach to sound synthesis (SDT) [32], and voice-driven synthesisers [21], as production means to support the embodiment of concepts into sound representations, depict a neutral to positive reception, mostly undermined by the steep learning curve, and the existing flaws of the research software. The potential of a software support to vocal sketching is reported in Q14.

Indeed, this interpretation is suggested by the participants’ comments on both the effectiveness and flaws of SEED. The embodied sound design tool was found “relatively fast and immediate as a high-level application”, especially for “live syncing and dubbing of sonic elements in video editing”. “The management of the SDT models seems quite intuitive”, and the voice-driven sound synthesis “quickly denatures the idea from the vocal character, thus allowing a timbral search in the development of the idea”. Major flaws concerned instead “the creation of melodies or some sound with harmony”, and “the development of complex sounds with specific behaviours”. On the sound implementation side, participants reported the “lack of a direct access to all families of sound models”, “the limited number of available sound models and synthesis techniques”, and in general “the need of sound libraries pliable to multiple solutions”.

Here, the relevant aspects are not the technical limitations and flaws, certainly representing an issue in terms of effective vocal control. What matters is the rationale underlying the development of any sound design tool aimed at empowering imagery through motor skills and control. Indeed, any creative freedom is necessarily bounded to the expression space afforded by the tool at hand, just as the

choice of a set of crayons and a certain paper would constrain a drawing act and give a material character to a visual sketch.

## 5. CONCLUSIONS

In Section 1, we posed the question about a terminological difference between designing sound and designing with sound. The question was instrumental to problematise the established view on sound design. A dual, apparently misleading, essence seems to emerge. The first is more traditional, and anchored in continuity with musical advances in computing technologies, in which designing sounds echoes implementing sounds with certain given aesthetic functional characteristics. The second perspective considers sound as a medium that can be a more or less appropriate host for representing abstract and concrete attributes of the artefact under conceptual inquiry. Designing with sound stresses the importance of embodied representations, internal or external, abstract or precise, etc., to cope with the complexity, ambiguity and incompleteness of design problems.

In this respect, an embodied approach to sound design and research may provide a key to understanding the problem of the perceived short distance between a sound design concept and the final product [23].

## 6. REFERENCES

- [1] N. Misdariis and A. Cera, “Knowledge in sound design: The silent electric vehicle? A relevant case study,” in *Proceedings of the Conference on Design and Semantics of Form and Movement-Sense and Sensitivity, DeSForM 2017*, InTech, 2017.
- [2] P. Susini, O. Houix, and N. Misdariis, “Sound design: an applied, experimental framework to study the perception of everyday sounds,” *The New Soundtrack*, vol. 4, no. 2, pp. 103–121, 2014.
- [3] E. Özcan and R. van Egmond, “Product sound design: An inter-disciplinary approach?,” in *Undisciplined! Design Research Society Conference*, 2009.
- [4] C. Erkut, S. Serafin, M. Hoby, and J. Särde, “Product sound design: Form, function, and experience,” in *Proceedings of the Audio Mostly 2015 on Interaction With Sound*, AM ’15, (New York, NY, USA), pp. 10:1–10:6, ACM, 2015.
- [5] D. Hug and M. Kemper, “From Foley to function: A pedagogical approach to sound design for novel interactions,” *Journal of Sonic Studies*, vol. 6, no. 1, 2014.
- [6] E. Özcan and M. Sonneveld, “Embodied explorations of sound and touch in conceptual design,” in *Design semantics of form and movement*, pp. 173–181, 2009.
- [7] W. Visser, “Design: one, but in different forms,” *Design studies*, vol. 30, no. 3, pp. 187–223, 2009.
- [8] M. Leman, *Embodied music cognition and mediation technology*. Cambridge, MA: MIT Press, 2008.

- [9] T. Van Rompay and G. Ludden, “Types of embodiment in design: The embodied foundations of meaning and affect in product design,” *International journal of design*, vol. 9, no. 1, 2015.
- [10] G. Lemaître, L. M. Heller, N. Navolio, and N. Zúñiga-Peñaranda, “Priming gestures with sounds,” *PLoS one*, vol. 10, no. 11, p. e0141791, 2015.
- [11] E. Pei, I. Campbell, and M. Evans, “A taxonomic classification of visual design representations used by industrial designers and engineering designers,” *The Design Journal*, vol. 14, no. 1, pp. 64–91, 2011.
- [12] G. Goldschmidt, “Manual sketching: Why is it still relevant?,” in *The Active Image: Architecture and Engineering in the Age of Modeling* (S. Ammon and R. Capdevila-Werning, eds.), pp. 77–97, Cham: Springer International Publishing, 2017.
- [13] J. Tholander, K. Karlgren, R. Ramberg, and P. Sökjer, “Where all the interaction is: Sketching in interaction design as an embodied practice,” in *Proceedings of the 7th ACM Conference on Designing Interactive Systems*, DIS ’08, (New York, NY, USA), pp. 445–454, ACM, 2008.
- [14] W. Visser and M. L. Maher, “The role of gesture in designing,” *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, vol. 25, no. 3, pp. 213–220, 2011.
- [15] D. Schleicher, P. Jones, and O. Kachur, “Bodystorming as embodied designing,” *Interactions*, vol. 17, no. 6, pp. 47–51, 2010.
- [16] S. Greenberg, S. Carpendale, N. Marquardt, and B. Buxton, *Sketching user experiences: The workbook*. Elsevier, 2011.
- [17] K. Franinović and S. Serafin, *Sonic interaction design*. Mit Press, 2013.
- [18] S. Delle Monache, S. Baldan, P. Boussard, A. Del Piccolo, C. Dendievel, G. Lemaître, H. Lachambre, O. Houix, and D. Rocchesso, “Interactive prototypes realised with the SkAT-VG tools,” deliverable of Project SkAT-VG, Iuav University of Venice, 2015.
- [19] S. Delle Monache and D. Rocchesso, “Physical computing II: Sketching sonic interactions,” in *Foundations in Sound Design for Embedded Media* (M. Filimowicz, ed.), New York: Routledge, 2018. *Forthcoming*.
- [20] S. Pauletto, “The voice delivers the threats, Foley delivers the punch: Embodied knowledge in Foley artistry,” in *The Routledge Companion to Screen Music and Sound* (M. Mera, R. Sadoff, and B. Winters, eds.), pp. 338–348, Routledge, 2017.
- [21] S. Delle Monache, D. Rocchesso, F. Bevilacqua, G. Lemaître, S. Baldan, and A. Cera, “Embodied sound design,” *International Journal of Human-Computer Studies*, vol. 118, pp. 47 – 59, 2018.
- [22] P. Helgason, “Sound initiation and source types in human imitations of sounds,” in *Proc. Fonetik*, (Stockholm, Sweden), pp. 83–89, 2014.
- [23] G. Lemaître, O. Houix, F. Voisin, N. Misdariis, and P. Susini, “Vocal imitations of non-vocal sounds,” *PLoS ONE*, vol. 11, no. 12, p. e0168167, 2016.
- [24] G. Lemaître, A. Jabbari, N. Misdariis, O. Houix, and P. Susini, “Vocal imitations of basic auditory features,” *The Journal of the Acoustical Society of America*, vol. 139, no. 1, pp. 290–300, 2016.
- [25] I. Ekman and M. Rinott, “Using vocal sketching for designing sonic interactions,” in *Proceedings of the 8th ACM Conference on Designing Interactive Systems*, DIS ’10, (New York, NY, USA), pp. 123–131, ACM, 2010.
- [26] S. Delle Monache and D. Rocchesso, “Understanding cooperative sound design through linkographic analysis,” in *Extending Interactivity - Proc. of the XXI CIM Colloquium on Music Informatics* (A. Terzaroli and A. Valle, eds.), pp. 25–32, 2016.
- [27] D. Rocchesso, *Explorations in sonic interaction design*. Logos Berlin, 2011.
- [28] A. Del Piccolo and D. Rocchesso, “Non-speech voice for sonic interaction: a catalogue,” *Journal on Multimodal User Interfaces*, vol. 11, no. 1, pp. 39–55, 2017.
- [29] P. Knees and K. Andersen, “Searching for audio by sketching mental images of sound: A brave new idea for audio retrieval in creative music production,” in *Proceedings of the 2016 ACM on International Conference on Multimedia Retrieval*, ICMR ’16, (New York, NY, USA), pp. 95–102, ACM, 2016.
- [30] D. Rocchesso, G. Lemaître, P. Susini, S. Ternström, and P. Boussard, “Sketching sound with voice and gesture,” *interactions*, vol. 22, no. 1, pp. 38–41, 2015.
- [31] D. Rocchesso, D. A. Mauro, and S. Delle Monache, “miMic: The microphone as a pencil,” in *Proceedings of the TEI ’16: Tenth International Conference on Tangible, Embedded, and Embodied Interaction*, TEI ’16, (New York, NY, USA), pp. 357–364, ACM, 2016.
- [32] S. Baldan, S. Delle Monache, and D. Rocchesso, “The sound design toolkit,” *SoftwareX*, vol. 6, pp. 255–260, 2017.
- [33] M. V. Mathews, “The digital computer as a musical instrument,” *Science*, vol. 142, no. 3592, pp. 553–557, 1963.
- [34] G. Schubert and F. Sallis, “Sketches and sketching,” in *A Handbook to Twentieth-Century Musical Sketches* (P. Hall and F. Sallis, eds.), pp. 5–16, Cambridge University Press, 2004.

# Embryo of a Quantum Vocal Theory of Sound

**Davide Rocchesso**

University of Palermo  
davide.rocchesso@unipa.it

**Maria Mannone**

University of Minnesota, alumna  
manno012@umn.edu

## ABSTRACT

Concepts and formalism from acoustics are often used to exemplify quantum mechanics. Conversely, quantum mechanics could be used to achieve a new perspective on acoustics, as shown by Gabor studies. Here, we focus in particular on the study of human voice, considered as a probe to investigate the world of sounds. We present a theoretical framework that is based on *observables* of vocal production, and on some *measurement apparati* that can be used both for analysis and synthesis. In analogy to the description of spin states of a particle, the quantum-mechanical formalism is used to describe the relations between the fundamental states associated with phonetic labels such as phonation, turbulence, and slow myoelastic vibrations. The intermingling of these states, and their temporal evolution, can still be interpreted in the Fourier/Gabor plane, and effective extractors can be implemented. This would constitute the basis for a Quantum Vocal Theory of sound, with implications in sound analysis and design.

## 1. INTRODUCTION

What are the fundamental elements of sound? What is the best framework for analyzing existing sonic realities and for expressing new sound concepts? These are long standing questions in sound physics, perception, and creation. In 1947, in a famous paper published in Nature [1], Dennis Gabor embraced the mathematics of quantum theory to shed light on subjective acoustics, thus laying the basis for sound analysis and synthesis based on acoustical quanta, or grains, or wavelets. The Fourier/Gabor framework for time-frequency representation of sound is widely used, although human acuity has been shown to beat the uncertainty limit [2] and cochlear filters [3] have been proposed to match human performance. Still, when we are imagining sound, or describing it to peers, we do not use the Fourier formalism.

### 1.1 Voice as Embodied Sound

Many researchers, in science, art, and philosophy, have been facing the problem of how to approach sound and its

representations [4, 5]. Should we represent sounds as they appear to the senses, by manipulating their proximal characteristics? Or should we rather look at potential sources, at physical systems that produce sound as a side effect of distal interactions? In this research path we assume that our body can help establishing bridges between distal (source-related) and proximal (sensory-related) representations, and we look at research findings in perception, production, and articulation of sounds [6]. Our embodied approach to sound [7] seeks to exploit knowledge in these areas, especially referring to human voice production as a form of embodied representation of sound.

When considering what people hear from the environment, it emerges that sounds are mostly perceived as belonging to categories of the physical world [8]. Research in sound perception has shown that listeners spontaneously create categories such as solid, electrical, gas, and liquid sounds, even though the sounds within these categories may be acoustically different [9]. However, when the task is to separate, distinguish, count, or compose sounds, the attention shifts from *sounding objects* to *auditory objects* [10] represented in the time-frequency plane. Tonal components, noise, and transients can be extracted from sound objects with Fourier-based techniques [11, 12, 13]. Low-frequency periodic phenomena are also perceptually very relevant and often come as trains of transients. The most prominent elements of the proximal signal may be selected by simplification and inversion of time-frequency representations. These auditory sketches [14], have been used to test the recognizability of imitations [15].

Vocal imitations can be more effective than verbalizations at representing and communicating sounds when these are difficult to describe with words [16]. This indicates that vocal imitations can be a useful tool for investigating sound perception, and shows that the voice is instrumental to embodied sound cognition. At a more fundamental level, research on non-speech vocalization is affecting the theories of language evolution [17], as it seems plausible that humans could have used iconic vocalizations to communicate with a large semantic spectrum, prior to the establishment of full-blown spoken languages. Experiments and sound design exercises [7] show that agreement in production corresponds to agreement in meaning interpretation, thus showing the effectiveness of teamwork in embodied sound creation. Converging evidences from behavioral and brain imaging studies give a firm basis to hypothesize a shared representation of sound in terms of motor (vocal) primitives [18].

Copyright: ©2018 Davide Rocchesso et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution License 3.0 Unported](https://creativecommons.org/licenses/by/3.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

In the recent EU-FET project SkAT-VG<sup>1</sup>, phoneticians have had an important role in identifying the most relevant components of non-speech voice productions [19]. They identified the broad categories of phonation (i.e., quasi periodic oscillations due to vocal fold vibrations), turbulence, slow myoelastic vibrations, and clicks, which can be extracted automatically from audio with time-frequency analysis and machine learning [20], and can be made to correspond to categories of sounds as they are perceived [15], and as they are produced in the physical world. Indeed, it has been argued that human utterances somehow mimic “nature’s phonemes” [21].

## 1.2 Quantum Frameworks

It was Dennis Gabor [1] who first adopted the mathematics of quantum mechanics to explain acoustic phenomena. In particular, he used operator methods to derive the time-frequency uncertainty relation and the (Gabor) function that satisfies minimal uncertainty. Time-scale representations [22] are more suitable to explain the perceptual decoupling of pitch and timbre, and operator methods can be used as well to derive the gammachirp function, which minimizes uncertainty in the time-scale domain [23]. Research in human and machine hearing [3] have been based on banks of elementary (filter) functions and these systems are at the core of many successful applications in the audio domain.

Despite its deep roots in the physics of the twentieth century, the sound field has not yet embraced the quantum signal processing framework [24] to seek practical solutions to sound scene representation, separation and analysis. A quantum approach to music cognition has not been proposed until very recently [25], when its explanatory power has been demonstrated to describe tonal attraction phenomena in terms of metaphorical forces. The theory of open quantum systems has been applied to music to describe the memory properties (non-Markovianity) of different scores [26]. In the image domain, on the other hand, it has been shown how the quantum framework can be effective to solve problems such as segmentation. For example, the separation of figures from background can be obtained by evolving a solution of the time-dependent Schrödinger equation [27], or by discretizing the time-independent Schrödinger equation [28]. An approach to signal manipulation based on the postulates of quantum mechanics can also potentially lead to a computational advantage when using Quantum Processing Units. Results in this direction are being reported for optimization problems [29].

## 1.3 Research Direction

In the proposed research path, sound is treated as a superposition of states, and the voice-based components (phonation, turbulence, slow myoelastic pulsations) are considered as observables to be represented as operators. The extractors of the fundamental components, i.e. the measurement apparatus, are implemented as signal-processing

modules that are available both for analysis and, as control knobs, for synthesis. The baseline is found in the results of the SkAT-VG project, which showed that vocal imitations are optimized representations of referent sounds, that emphasize those features that are important for identification. A large collection of audiovisual recordings of vocal and gestural imitations<sup>2</sup> offers the opportunity to further enquire how people perceive, represent, and communicate about sounds.

A first assumption underlying this research approach, largely justified by prior art and experiences, is that articulatory primitives used to describe vocal utterances are effective as high-level descriptors of sound in general. This assumption leads naturally to an embodied approach to sound representation, analysis, and synthesis. A second assumption is that the mathematics of quantum mechanics, relying on linear operators in Hilbert spaces, offers a formalism that is suitable to describe the objects composing auditory scenes and their evolution in time. The latter assumption is more adventurous, as this path has not been taken in audio signal processing yet. However, the results coming from neighboring fields (music cognition, image processing) encourage us to explore this direction, and to aim at improved techniques for sound analysis and synthesis.

## 2. SKETCH OF A QUANTUM VOCAL THEORY

An embryonic theory of sound based on the postulates of quantum mechanics, and using high-level vocal descriptors of sound, can be sketched as follows. Let  $\sigma$  be a vector operator that provides information about the phonetic elements along a specific *direction* of measurement. Phonation, for example, may be represented by  $\sigma_z$ , with eigenstates representing an upper and a lower pitch. Similarly, the turbulence component may be represented by  $\sigma_x$ , with eigenstates representing turbulence of two different distributions. A measurement of turbulence prepares the *system* in one of two eigenstates for operator  $\sigma_x$ , and a successive measurement of phonation would find a superposition and get equal probabilities for the two eigenstates of  $\sigma_z$ . The two operators  $\sigma_z$  and  $\sigma_x$  may also be made to correspond to the two components of the classic sines+noise model used in audio signal processing. If we add transients/clicks as a third measurement direction (as in the sines + noise + transients model [12]) we can claim that there is no sound state for which the expectation value of the three components is zero: a sort of spin polarization principle as found in quantum mechanics. The evolution of state vectors in time is unitary, and regulated by a time-dependent Schrödinger equation, with a suitably chosen Hamiltonian. The eigenvectors of the Hamiltonian allow to expand any state vector in that basis, and to compute the time evolution of such expansion. A pair of components can be simultaneously measured only if they commute. If they don’t, an uncertainty principle can be derived, as it was done for time-frequency and time-scale representations [1, 23]. The theory can be extended to cover multiple sources, and the resulting mixed

<sup>1</sup> [www.skatvg.eu](http://www.skatvg.eu)

<sup>2</sup> <https://www.ircam.fr/projects/blog/multimodal-database-of-vocal-and-gestural-imitations-elicited-by-sounds/>

states can be described via density matrices, whose time evolution can also be computed if a Hamiltonian operator is properly defined. The effect of disturbances and deviations can be accounted for by introducing relaxation time, to model return to equilibrium. This sketch of a Quantum Vocal Theory needs to be developed and formally laid down.

### 3. THE PHON FORMALISM

Consider a 3d space with the orthogonal axes

**z** : phonation, with different pitches;

**x** : turbulence, with noises of different frequency distributions;

**y** : myoelastic, slow pulsations with different tempos.

The phon operator  $\sigma$  is a 3-vector operator that provides information about the phonetic elements in a specific direction of the 3d phonetic space.

In this section we present the phon formalism, obtained by direct analogy with the single spin, as presented in accessible presentations of quantum mechanics [30]. We use standard Dirac notation.

#### 3.1 Measurement along z

A measurement along the z axis is performed according to the quantum-mechanics principles:

1. Each component of  $\sigma$  is represented by a linear operator;
2. The eigenvectors of  $\sigma_z$  are  $|u\rangle$  and  $|d\rangle$ , corresponding to pitch up and pitch down, with eigenvalues  $+1$  and  $-1$ , respectively:

$$(a) \sigma_z |u\rangle = |u\rangle$$

$$(b) \sigma_z |d\rangle = -|d\rangle$$

3. The eigenstates of operator  $\sigma_z$ ,  $|u\rangle$  and  $|d\rangle$ , are orthogonal:  $\langle u|d\rangle = 0$ ;

The eigenstates can be represented as column vectors:

$$|u\rangle = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \text{ and } |d\rangle = \begin{bmatrix} 0 \\ 1 \end{bmatrix},$$

and the operator  $\sigma_z$  as a square  $2 \times 2$  matrix. Due to principle 2, we have

$$\sigma_z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}. \quad (1)$$

#### 3.2 Preparation along x

The eigenstates of the operator  $\sigma_x$  are  $|r\rangle$  and  $|l\rangle$ , corresponding to turbulences having different spectral distributions, one with the rightmost centroid and the other with the leftmost centroid. The respective eigenvalues are  $+1$  and  $-1$ , so that

$$(a) \sigma_x |r\rangle = |r\rangle$$

$$(b) \sigma_x |l\rangle = -|l\rangle.$$

If the phon is prepared  $|r\rangle$  (turbulent) and then the measurement apparatus is set to measure  $\sigma_z$ , there will be equal probabilities for  $|u\rangle$  or  $|d\rangle$  phonation as an outcome. Essentially, we are measuring what kind of phonation is in a pure turbulent state. This measurement property is satisfied if

$$|r\rangle = \frac{1}{\sqrt{2}} |u\rangle + \frac{1}{\sqrt{2}} |d\rangle. \quad (2)$$

In fact, any state  $|A\rangle$  can be expressed as  $|A\rangle = \alpha_u |u\rangle + \alpha_d |d\rangle$ , where  $\alpha_u = \langle u|A\rangle$ , and  $\alpha_d = \langle d|A\rangle$ . The probability to measure pitch up is  $P_u = \langle A|u\rangle \langle u|A\rangle = \alpha_u^* \alpha_u$ , and the probability to measure pitch down is  $P_d = \langle A|d\rangle \langle d|A\rangle = \alpha_d^* \alpha_d$  (Born rule).

Likewise, if the phon is prepared  $|l\rangle$  and then the measurement apparatus is set to measure  $\sigma_z$ , there will be equal probabilities for  $|u\rangle$  or  $|d\rangle$  phonation as an outcome. This measurement property is satisfied if

$$|l\rangle = \frac{1}{\sqrt{2}} |u\rangle - \frac{1}{\sqrt{2}} |d\rangle, \quad (3)$$

which is orthogonal to the linear combination (2). In vector

form, we have:  $|r\rangle = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$  and  $|l\rangle = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix}$ ,

and

$$\sigma_x = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}. \quad (4)$$

#### 3.3 Preparation along y

The eigenstates of the operator  $\sigma_y$  are  $|f\rangle$  and  $|s\rangle$ , corresponding to slow myoelastic pulsations, one faster and one slower<sup>3</sup>, with eigenvalues  $+1$  and  $-1$ , so that

$$(a) \sigma_y |f\rangle = |f\rangle$$

$$(b) \sigma_y |s\rangle = -|s\rangle$$

If the phon is prepared  $|f\rangle$  (pulsating) and then the measurement apparatus is set to measure  $\sigma_z$ , there will be equal probabilities for  $|u\rangle$  or  $|d\rangle$  phonation as an outcome. Essentially, we are measuring what kind of phonation is in a slow myoelastic pulsations. This measurement property is satisfied if

$$|f\rangle = \frac{1}{\sqrt{2}} |u\rangle + \frac{i}{\sqrt{2}} |d\rangle, \quad (5)$$

where  $i$  is the imaginary unit.

Likewise, if the phon is prepared  $|s\rangle$ , we can express this state as

$$|s\rangle = \frac{1}{\sqrt{2}} |u\rangle - \frac{i}{\sqrt{2}} |d\rangle, \quad (6)$$

which is orthogonal to the linear combination (5). In vector

form, we have:  $|f\rangle = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{i}{\sqrt{2}} \end{bmatrix}$  and  $|s\rangle = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{i}{\sqrt{2}} \end{bmatrix}$ ,

and

$$\sigma_y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}. \quad (7)$$

The matrices (1), (4), and (7) are called the Pauli matrices and, together with the identity matrix, these are the quaternions.

<sup>3</sup> In describing the spin eigenstates, the symbols  $|i\rangle$  and  $|o\rangle$  are often used, to denote the in-out direction.

### 3.4 Measurement along an arbitrary direction

Orienting the measurement apparatus along an arbitrary direction  $\bar{n} = [n_x, n_y, n_z]'$  means taking a weighted mixture:

$$\begin{aligned} \sigma_n &= \bar{\sigma} \cdot \bar{n} = \sigma_x n_x + \sigma_y n_y + \sigma_z n_z = \\ &= \begin{bmatrix} n_z & n_x - i n_y \\ n_x + i n_y & -n_z \end{bmatrix}. \end{aligned} \quad (8)$$

#### 3.4.1 Example: Harmonic plus Noise model

A measurement performed by means of a Harmonic plus Noise model [11] would lie in the phonation-turbulence plane ( $n_z = \cos \theta, n_x = \sin \theta, n_y = 0$ ), so that

$$\sigma_n = \begin{bmatrix} \cos \theta & \sin \theta \\ \sin \theta & -\cos \theta \end{bmatrix} \quad (9)$$

The eigenstate for eigenvalue  $+1$  is  $|\lambda_1\rangle = \begin{bmatrix} \cos \theta/2 \\ \sin \theta/2 \end{bmatrix}$ , the eigenstate for eigenvalue  $-1$  is  $|\lambda_{-1}\rangle = \begin{bmatrix} -\sin \theta/2 \\ \cos \theta/2 \end{bmatrix}$ , and the two are orthogonal. Suppose we prepare the phon to pitch-up  $|u\rangle$ . If we rotate the measurement system along  $\bar{n}$ , the probability to measure  $\sigma_n = +1$  is (by Born rule)

$$P(+1) = |\langle u | \lambda_1 \rangle|^2 = \cos^2 \theta/2, \quad (10)$$

and the probability to measure  $\sigma_n = -1$  is

$$P(-1) = |\langle u | \lambda_{-1} \rangle|^2 = \sin^2 \theta/2. \quad (11)$$

The expectation value of measurement is therefore

$$\begin{aligned} \langle \sigma_n \rangle &= \sum_i \lambda_i P(\lambda_i) \\ &= (+1) \cos^2 \theta/2 + (-1) \sin^2 \theta/2 = \cos \theta \end{aligned} \quad (12)$$

#### 3.4.2 Rotate to measure

What does it mean to rotate a measurement apparatus to measure a property? Assume we have a machine that separates harmonics from noise from (trains of) transients, and that can discriminate between two different pitches, noise distributions, and tempos. Essentially, the machine receives a sound and returns three numbers  $\{\text{ph}, \text{tu}, \text{my}\} \in [-1, 1]$ . If  $\text{ph} > 0$  the result will be  $|u\rangle$ , and if  $\text{ph} < 0$  the result will be  $|d\rangle$ . If  $\text{tu} > 0$  the result will be  $|r\rangle$ , and if  $\text{tu} < 0$  the result will be  $|l\rangle$ . If  $\text{my} > 0$  the result will be  $|f\rangle$ , and if  $\text{my} < 0$  the result will be  $|s\rangle$ . These three outputs correspond to rotating the measurement apparatus along each of the main axes. Rotating it along an arbitrary direction means taking a weighted mixture of the three outcomes.

### 3.5 Time evolution

In quantum mechanics the evolution of state vectors in time

$$|\psi(t)\rangle = \mathbf{U}(t) |\psi(0)\rangle \quad (13)$$

is governed by the operator  $\mathbf{U}$ , which is unitary, i.e.,  $\mathbf{U}^\dagger \mathbf{U} = \mathbf{I}$ . Taken a small time increment  $\epsilon$ , continuity of the time-development operator gives it the form

$$\mathbf{U}(\epsilon) = \mathbf{I} - i\epsilon \mathbf{H}, \quad (14)$$

with  $\mathbf{H}$  being the quantum Hamiltonian (Hermitian) operator.  $\mathbf{H}$  is an observable and its eigenvalues are the values that would result from measuring the energy of a quantum system. From (14) it turns out that a state vector changes in time according to the time-dependent Schrödinger equation<sup>4</sup>

$$\frac{\partial |\psi\rangle}{\partial t} = -i\mathbf{H} |\psi\rangle. \quad (15)$$

Any observable  $\mathbf{L}$  has an expectation value  $\langle \mathbf{L} \rangle$  that evolves according to

$$\frac{\partial \langle \mathbf{L} \rangle}{\partial t} = -i \langle [\mathbf{L}, \mathbf{H}] \rangle, \quad (16)$$

where  $[\mathbf{L}, \mathbf{H}] = \mathbf{L}\mathbf{H} - \mathbf{H}\mathbf{L}$  is the commutator of  $\mathbf{L}$  with  $\mathbf{H}$ .

#### 3.5.1 Phon in utterance field

Similarly to a spin in a magnetic field, when a phon is part of an utterance, it has an energy that depends on its orientation. We can think about it as if it was subject to restoring forces, and its quantum Hamiltonian is

$$\mathbf{H} \propto \bar{\sigma} \cdot \bar{B} = \sigma_x B_x + \sigma_y B_y + \sigma_z B_z, \quad (17)$$

where the components of the field  $\bar{B}$  are named in analogy with the magnetic field.

Consider the case of potential energy only along  $z$ :

$$H = \frac{\omega}{2} \sigma_z. \quad (18)$$

To find how the expectation value of the phon varies in time, we expand the observable  $\mathbf{L}$  in (16) in its components to get

$$\begin{aligned} \langle \dot{\sigma}_x \rangle &= -i \langle [\sigma_x, \mathbf{H}] \rangle = -\omega \langle \sigma_y \rangle \\ \langle \dot{\sigma}_y \rangle &= -i \langle [\sigma_y, \mathbf{H}] \rangle = \omega \langle \sigma_x \rangle \\ \langle \dot{\sigma}_z \rangle &= -i \langle [\sigma_z, \mathbf{H}] \rangle = 0, \end{aligned} \quad (19)$$

which means that the expectation values of  $\sigma_x$  and  $\sigma_y$  are subject to temporal precession around  $z$  at angular velocity  $\omega$ . The expectation value of  $\sigma_z$  steadily keeps the pitch if there is no potential energy along turbulence and slow myoelastic vibration.

A potential energy along all three axes can be expressed as

$$\mathbf{H} = \frac{\omega}{2} \bar{\sigma} \cdot \bar{n} = \frac{\omega}{2} \begin{bmatrix} n_z & n_x - i n_y \\ n_x + i n_y & -n_z \end{bmatrix}, \quad (20)$$

whose energy eigenvalues are  $E_j = \pm 1$ , with energy eigenvectors  $|E_j\rangle$ .

<sup>4</sup> We do not need physical dimensional consistency here, so we drop Planck's constant.

An initial state vector (phon)  $|\psi(0)\rangle$  can be expanded as

$$|\psi(0)\rangle = \sum_j \alpha_j(0) |E_j\rangle, \quad (21)$$

where  $\alpha_j(0) = \langle E_j | \psi(0) \rangle$ , and the time evolution of state turns out to be

$$|\psi(t)\rangle = \sum_j \alpha_j(t) |E_j\rangle = \sum_j \alpha_j(0) e^{-iE_j t} |E_j\rangle. \quad (22)$$

### 3.6 Uncertainty

If we measure two observables  $L$  and  $M$  (in a single experiment) simultaneously, quantum mechanics prescribes that the system is left in a simultaneous eigenvector of the observables only if  $L$  and  $M$  commute, i.e.  $[L, M] = 0$ . Measurement operators along different axes do not commute. For example,  $[\sigma_x, \sigma_y] = 2i\sigma_z$ , and therefore phonation and turbulence can not be simultaneously measured with certainty.

The uncertainty principle, based on Cauchy-Schwarz inequality in complex vector spaces, prescribes that the product of the two uncertainties is at least as large as half the magnitude of the commutator:

$$\Delta L \Delta M \geq \frac{1}{2} |\langle \psi | [L, M] | \psi \rangle| \quad (23)$$

Equation (23) expresses the uncertainty principle.

If  $L = t$  is the time operator and  $S = -i \frac{d}{dt}$  is the frequency operator, and these are applied to the complex oscillator  $Ae^{i\omega t}$ , the time-frequency uncertainty principle results, and uncertainty is minimized by the Gabor function. Starting from the scale operator, the gammachirp function can be derived [23].

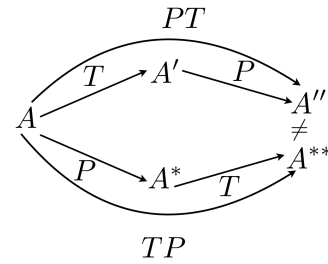
## 4. WHAT TO DO WITH ALL THIS

Assuming that a vocal description of sound follows the postulates of quantum mechanics, the question is how to take advantage of the quantum formalism. How can we practically connect theoretical ideas with audio signal processing?

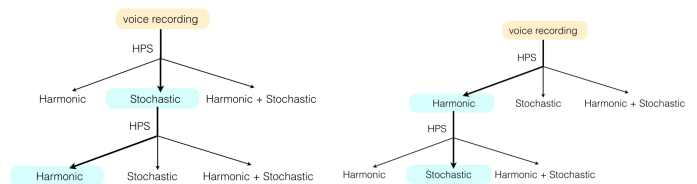
### 4.1 Non-commutativity and autostates

We expect that measurement operators along different axes do not commute: this is the case, for example, of measurements of phonation and turbulence. Let  $A$  be an audio sample. The measurement of turbulence by the operator  $T$  leads to  $T(A) = A'$ . A successive measurement of phonation by the operator  $P$  gives  $P(A') = A''$ , thus  $P(A') = PT(A) = A''$ . If we perform the measurements in the opposite order, with phonation first and turbulence later, we obtain  $TP(A) = T(A^*) = A^{**}$ . We expect that  $[T, P] \neq 0$ , and thus, that  $A^{**} \neq A''$ . The diagram in figure 1 shows non-commutativity in the style of category theory.

Measurements of phonation and turbulence can be actually performed using the Harmonic plus Stochastic (HPS) model [11]. The order of operations is visually described in figure 2. The measurement of phonation is performed



**Figure 1.** A non-commutative diagram (in the style of category theory), representing the non-commutativity of measurements of phonation ( $P$ ) and turbulence ( $T$ ) on audio  $A$ .

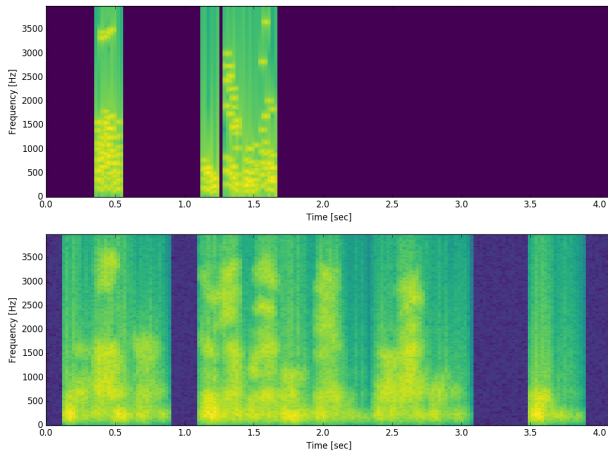


**Figure 2.** On the left, a voice is analyzed via the HPS model. Then, the stochastic part is submitted to a new analysis. In this way, a measurement of phonation follows a measurement of turbulence. On the right, the measurement of turbulence follows a measurement of phonation.

through the extraction of the harmonic component in the HPS model, while the measurement of turbulence is performed through the extraction of the stochastic component with the same model. The spectrograms for  $A''$  and  $A^{**}$  in Figure 3 show the results of such two sequences of analyses on a segment of female speech<sup>5</sup>, confirming that the commutator  $[T, P]$  is non-zero.

Essentially, if we adopt the HPS model and skip the final step of addition and inverse transformation, we are left with something that is conceptually equivalent to a quantum *destructive measure*. Let  $St$  be the filter that extracts the stochastic part from a signal. As figure 4 shows, the spectrogram of  $St(x)$  is visibly different from the spectrogram of  $x$ . Conversely, if we apply  $St$  once more, we get a spectrum that does not change much:  $St^2(x) = St(St(x)) \sim St(x)$ . If we transform back from the second and third spectrograms of figure 4, we get sounds that are very close to each other. In fact, ideally,  $St^2(x) = St(x)$ . It means that, after a measure of the non-harmonic component of some signal, the output-signal can be considered as an *autostate*. If we perform the measure again and again, we still get the same result. Such a measure operation provokes the collapse of a hypothetical underlying wave function, which is originally a superposition of states, and is reduced to a single state upon measurement. The importance of the autostates in this framework is connected with the concept

<sup>5</sup> <https://freesound.org/s/317745/>. Hann window of 2048 samples, FFT of 4096 samples, hop size of 1024 samples.



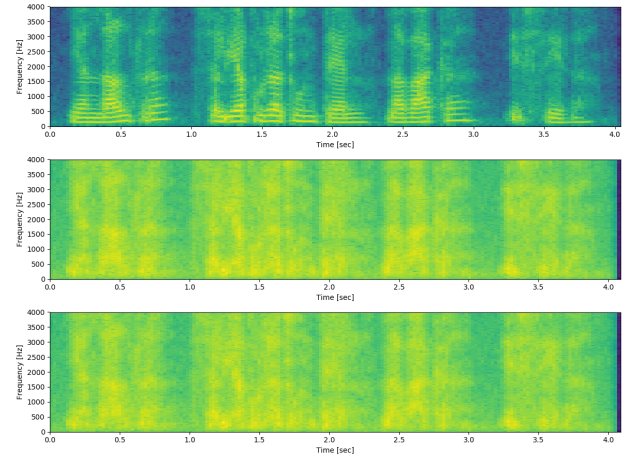
**Figure 3.** On the top, the spectrogram corresponding to a measurement of phonation  $P$  following a measurement of turbulence  $T$ , leading to  $PT(A) = A''$ . On the bottom, the spectrogram corresponding to a measurement of turbulence  $T$  following a measurement of phonation  $P$ , leading to  $TP(A) = A^{**}$ .

of quantum measures, which may become practically feasible through a set of audio-signal analysis tools.

We can define other filters that extract more specific information: for example, a glissando filter would extract glissando passages within a sound signal, giving as output the inverse transform of the filtered spectrum, that is, a filtered signal where we can just hear the glissando effect and nothing else. Re-performing the glissando-measure on that sound, we still get the same sound: the new sound signal is an autostate of glissando. A suitable collection of filters that act both on the original sounds as well as on their vocal imitations can produce simplified spectra, easier to compare. These may give hints on how information are extracted by voice, that permit to recognize the original sound source. This would confirm the importance of human voice as a probe to investigate the world of sounds, and Quantum Vocal Theory as a bridge between quantum physics, acoustics, and cognition, with possible further bridges to multisensory perception and interaction. If we consider that voice can imitate not only sounds, but also movements and, sometimes, even visual shapes through crossmodal correspondences [31], new fascinating scenarios open up for investigation.

## 4.2 Time evolution

We know that time evolution of states is governed by the unitary transformation (13) and by the Schrödinger equation (15). A measurement is represented by an operator that acts on the state and that causes its collapse onto one of its eigenvectors. The system remains in a superposition of states until a measurement occurs, whose outcome is inherently probabilistic. The states change under the effect of external forces, which therefore determine the change in the probabilities associated with each state. It is a Hamiltonian such as (20) that controls the evolution process. We can think of the vocal Hamiltonian as containing the po-



**Figure 4.** Top: spectrum of the original sound signal (a female speech), Center: the stochastic component, derived from harmonic plus stochastic analysis (HPS), as the effect of a destructive measure, and Bottom: the stochastic component of the stochastic component itself. The last two spectra are very close.

tential energies due to restoring forces that are contained in the utterance at a particular moment in time. We could even think of a formal composition of forces acting at different levels, and describe them through a nested categorical depiction.

Similarly to what has been done by Youssry et al. [27], the Hamiltonian can be chosen to be time-dependent yet commutative (i.e.,  $[\mathbf{H}(0), \mathbf{H}(t)] = \mathbf{H}(0)\mathbf{H}(t) - \mathbf{H}(t)\mathbf{H}(0) = 0$ ), so that a closed-form solution to state evolution can be obtained. A time-independent Hamiltonian such as the one leading to (22) would not be very useful, both because forces indeed change continuously and because this would lead to oscillatory solution. Instead, with a time-varying commutative Hamiltonian the time evolution can be expressed as

$$|\psi(t)\rangle = e^{-i \int_0^t H(\tau) d\tau} |\psi(0)\rangle. \quad (24)$$

A simple choice is that of a Hamiltonian such as

$$H(t) = g(t)\mathbf{S}, \quad (25)$$

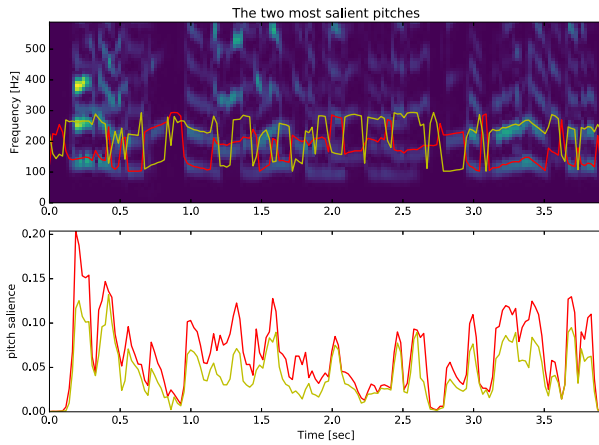
with  $\mathbf{S}$  a time-independent Hermitian matrix. A function  $g(t)$  that ensures convergence of the integral in (24) is the damping

$$g(t) = e^{-t}. \quad (26)$$

In an audio application, we can consider a slice of time and the initial and final states for that slice. We should look for a Hamiltonian that leads to the evolution of the initial state into the final state. In image segmentation [27], where time is used to let each pixel evolve to a final foreground-background assignment, the Hamiltonian is chosen to be

$$H = e^{-t} f(\mathbf{x}) \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \quad (27)$$

and  $f(\cdot)$  is a two-valued function of a feature vector  $\mathbf{x}$  that contains information about a neighborhood of the pixel.



**Figure 5.** Extraction of the two most salient pitches from a mixture of a male voice and a female voice

Such function is learned from an example image with a given ground truth. In audio we may do something similar and learn from examples of transformations: phonation to phonation, with or without pitch crossing; phonation to turbulence; phonation to myoelastic, etc. We may also add a coefficient to the exponent in (26), to govern the rapidity of transformation. As opposed to image processing, time is our playground.

The matrix  $\mathbf{S}$  can be set to assume the structure (20), and the components of potential energy found in an utterance field can be extracted as audio features. For example, pitch salience can be extracted from time-frequency analysis [32] and used as  $n_z$  component for the Hamiltonian. Figure 5 shows the two most salient pitches, automatically extracted from a mixture of male and female voice<sup>6</sup>. Frequent up-down jumps are evident, and they make difficult to track a single voice. Quantum measurement induces state collapse to  $|u\rangle$  or  $|d\rangle$  and, from that state, evolution can be governed by (24). In this way, it should be possible to mimic human figure-ground attention [33], and follow each individual voice.

## 5. CONCLUSION AND FURTHER RESEARCH

We presented an early attempt at applying the fundamental quantum formalism to the description of human voice, and to sounds in general through voice-based basic elements. Such a theoretical research can have several practical implications, due to the importance of voice as a probe to investigate the world of sounds in general. Also, a quantum vocal theory enhances the role of quantum mechanics and of the underlying mathematics as a connecting tool between different areas of human knowledge. By flipping the wicked problem of finding intuitive interpretations of quantum mechanics, the proposed approach aims to use quantum mechanics to interpret something that we have embodied, intuitive knowledge of. The new theory can lead to the development of voice-based tools for sound analysis and synthesis, for research and communication

<sup>6</sup> <https://freesound.org/s/431595/>

understanding. Finally, given the importance of vocal imitations in the framework of improvised music and sound art, they can constitute a probe also for studying and creating instrumental music, and for conceiving new forms of sonic expression, also through manual and body gestures, which enable connections to the visual and performing arts.

## 6. REFERENCES

- [1] D. Gabor, “Acoustical quanta and the theory of hearing,” *Nature*, vol. 159, no. 4044, p. 591, 1947.
- [2] J. N. Oppenheim and M. O. Magnasco, “Human time-frequency acuity beats the fourier uncertainty principle,” *Phys. Rev. Lett.*, vol. 110, p. 044301, Jan 2013.
- [3] R. F. Lyon, *Human and machine hearing*. Cambridge University Press, 2002.
- [4] G. De Poli, A. Piccialli, and C. Roads, *Representations of musical signals*. MIT Press, 1991.
- [5] D. Roden, “Sonic art and the nature of sonic events,” *Review of Philosophy and Psychology*, vol. 1, no. 1, pp. 141–156, 2010.
- [6] M. Leman, *Embodied music cognition and mediation technology*. MIT Press, 2008.
- [7] S. D. Monache, D. Rocchesso, F. Bevilacqua, G. Lemaitre, S. Baldan, and A. Cera, “Embodied sound design,” *International Journal of Human-Computer Studies*, vol. 118, pp. 47 – 59, 2018.
- [8] W. W. Gaver, “How do we hear in the world? Explorations in ecological acoustics,” *Ecological Psychology*, vol. 5, no. 4, pp. 285–313, 1993.
- [9] O. Houix, G. Lemaitre, N. Misdariis, P. Susini, and I. Urdapilleta, “A lexical analysis of environmental sound categories,” *Journal of Experimental Psychology: Applied*, vol. 18, no. 1, p. 52, 2012.
- [10] M. Kubovy and M. Schutz, “Audio-visual objects,” *Review of Philosophy and Psychology*, vol. 1, no. 1, pp. 41–61, 2010.
- [11] J. Bonada, X. Serra, X. Amatriain, and A. Loscos, “Spectral processing,” in *DAFX: Digital Audio Effects* (U. Zölzer, ed.), pp. 393–445, John Wiley & Sons, Ltd, 2011.
- [12] T. S. Verma, S. N. Levine, and T. H. Meng, “Transient Modeling Synthesis: a flexible analysis/synthesis tool for transient signals,” in *Proceedings of the International Computer Music Conference*, pp. 48–51, 1997.
- [13] R. Fg, A. Niedermeier, J. Driedger, S. Disch, and M. Miller, “Harmonic-percussive-residual sound separation using the structure tensor on spectrograms,” in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 445–449, March 2016.

- [14] V. Isnard, M. Taffou, I. Viaud-Delmon, and C. Suied, "Auditory sketches: very sparse representations of sounds are still recognizable," *PLoS one*, vol. 11, no. 3, 2016.
- [15] G. Lemaitre, A. Jabbari, N. Misdariis, O. Houix, and P. Susini, "Vocal imitations of basic auditory features," *The Journal of the Acoustical Society of America*, vol. 139, no. 1, pp. 290–300, 2016.
- [16] G. Lemaitre and D. Rocchesso, "On the effectiveness of vocal imitations and verbal descriptions of sounds," *The Journal of the Acoustical Society of America*, vol. 135, no. 2, pp. 862–873, 2014.
- [17] M. Perlman and G. Lupyan, "People can create iconic vocalizations to communicate various meanings to naïve listeners," *Scientific reports*, vol. 8, no. 1, p. 2634, 2018.
- [18] Z. Wallmark, M. Iacoboni, C. Deblieck, and R. A. Kendall, "Embodied Listening and Timbre: Perceptual, Acoustical, and Neural Correlates," *Music Perception: An Interdisciplinary Journal*, vol. 35, no. 3, pp. 332–363, 2018.
- [19] P. Helgason, "Sound initiation and source types in human imitations of sounds," in *Proceedings of FONETIK 2014*, pp. 83–88, 2014.
- [20] E. Marchetto and G. Peeters, "Automatic recognition of sound categories from their vocal imitation using audio primitives automatically derived by SI-PLCA and HMM," in *Proceedings of the International Symposium on Computer Music Multidisciplinary Research*, 2017.
- [21] M. Changizi, *Harnessed: How language and music mimicked nature and transformed ape to man*. Ben-Bella Books, Inc., 2011.
- [22] A. De Sena and D. Rocchesso, "A fast Mellin and scale transform," *EURASIP Journal on Applied Signal Processing*, no. Article ID 8917, 2007.
- [23] T. Irino and R. D. Patterson, "A time-domain, level-dependent auditory filter: The gammachirp," *The Journal of the Acoustical Society of America*, vol. 101, no. 1, pp. 412–419, 1997.
- [24] Y. C. Eldar and A. V. Oppenheim, "Quantum signal processing," *IEEE Signal Processing Magazine*, vol. 19, no. 6, pp. 12–32, 2002.
- [25] P. beim Graben and R. Blutner, "Quantum approaches to music cognition," *arXiv:1712.07417*, 2018.
- [26] M. Mannone and G. Compagno, "Characterization of the degree of Musical non-Markovianity," *arXiv:1306.0229*, 2013.
- [27] A. Youssry, A. El-Rafei, and S. Elramly, "A quantum mechanics-based framework for image processing and its application to image segmentation," *Quantum Information Processing*, vol. 14, no. 10, pp. 3613–3638, 2015.
- [28] Ç. Aytekin, E. C. Ozan, S. Kiranyaz, and M. Gabbouj, "Extended quantum cuts for unsupervised salient object extraction," *Multimedia Tools and Applications*, vol. 76, no. 8, pp. 10443–10463, 2017.
- [29] F. Neukart, G. Compostella, C. Seidel, D. von Dollen, S. Yarkoni, and B. Parney, "Traffic flow optimization using a quantum annealer," *Frontiers in ICT*, vol. 4, p. 29, 2017.
- [30] L. Susskind and A. Friedman, *Quantum Mechanics: The Theoretical Minimum*. UK: Penguin Books, 2015.
- [31] C. Spence, "Crossmodal correspondences: a tutorial review," *Attention, Perception, and Psychophysics*, vol. 73, no. 4, pp. 971–995, 2011.
- [32] J. Salamon and E. Gomez, "Melody extraction from polyphonic music signals using pitch contour characteristics," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, pp. 1759–1770, Aug 2012.
- [33] E. Bigand, S. McAdams, and S. Forêt, "Divided attention in music," *International Journal of Psychology*, vol. 35, no. 6, pp. 270–278, 2000.

# SULLA MUSICA MECCANICA

## A PARTIRE DA JOHN PHILIP SOUSA

**Antonio Mastrogiacomo**  
Conservatorio "N. Sala" Benevento  
ant.mastrogiacomo@gmail.com

### ABSTRACT

Se la possibilità di registrare, riascoltare e manipolare il suono ha rappresentato un momento delicato della storia della percezione nei tempi misurati della vita quotidiana, allora la diffusione incontrollata degli apparati – termine con cui intendiamo i media tecnici della modernità, sottolineandone il carattere preparato e quindi intenzionale, artificiale, convenzionale – si è radicata nell'ampio orizzonte del medium, termine convocato già nel saggio sulla riproducibilità tecnica per denominare il modo secondo cui si organizza la percezione umana [1]. E se alla radice della teoria dei media risiede un'idea secondo cui l'esperienza sensibile non può essere considerata in termini sovrastorici, allora le parole di Sousa hanno il merito di storicizzare questo delicato momento di passaggio nella misurazione dei tempi e delle tecnologie della produzione e dell'ascolto.

A partire dal saggio *La minaccia della musica meccanica* (1906) di John Philip Sousa [2], il contributo proposto discute funzione e fruizione della musica meccanica e meccanicizzata; in riferimento all'esperienza del suono nei limiti della società di massa, altri cinque diversi autori (Adorno, Benjamin, Edison, McLuhan e Moholy-Nagy) sono posti in dialogo con il saggio del compositore e direttore di banda statunitense.

### 1. INTRODUZIONE

John Philip Sousa pubblica nel 1906 un saggio dal titolo *The Menace of Mechanical Music* nell'*Appleton's Magazine* [2]. Si tratta del saggio cui allude Marshall McLuhan negli *Strumenti del comunicare* [3], al capitolo 28 – quello dedicato al grammofo.

*“Un uccello dal canto genuino rischia la sua estinzione sotto i colpi del passerotto inglese in grado di moltiplicare sé stesso”* [2].<sup>1</sup>

Copyright: © 2018 Antonio Mastrogiacomo. This is an open-access article distributed under the terms of the [Creative Commons Attribution License 3.0 Unported](https://creativecommons.org/licenses/by/3.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

<sup>1</sup> Traduzioni del testo di Sousa a cura dell'autore.

John Philip Sousa presenta la musica meccanica con queste parole: la vede come una minaccia per il petto della nazione, un pericolo imminente per l'American musical art da un lato, dall'altro un pericolo per i compositori per via del vuoto legislativo – si impegna a tal punto da essere tra i fondatori dell'ASCAP<sup>2</sup> (American Society of Composers, Authors and Publishers), 8 anni più tardi.

*“Si abbatte sul paese con la velocità di una moda transitoria nel gergo o nel panama, nei pianti sulle guerre politiche o nei romanzi popolari. Si tratta di un dispositivo meccanico che ci canta una canzone o ci suona il pianoforte, in sostituzione delle abilità, dell'intelligenza e dell'animo umano”* [2].

Tra la prima versione del saggio sull'opera d'arte di Walter Benjamin<sup>3</sup> - datata 1936 - e il testo di Sousa campeggia nel titolo il destino comune imposto dalla meccanicità all'opera d'arte. Sousa infarcisce il saggio di lucidissime osservazioni, quasi a profetizzare condizioni intraviste all'alba della musica meccanica. Lontano dalla gravezza del risentito Adorno, un forte trasporto emotivo traghetta infatti il lettore in una scrittura quasi partigiana.

*“Su una questione cui tengo profondamente e che considero di vasta portata sono disposto a essere considerato un allarmista, una questione su cui di certo sono influenzato parzialmente da interessi personali, nonché dall'imminente pericolo per l'American musical art. Prevedo un marcato deterioramento nei gusti musicali americani, una interruzione per lo sviluppo musicale del Paese e uno spazio per altre ferite per la musica nelle sue manifestazioni artistiche, in virtù della moltiplicazione di varie macchine riproduttrici”* [2].

Eppure lo strumento meccanico prima, elettronico in seguito ha da servire diverse posizioni teso com'è tra la

<sup>2</sup> Fondata il 13 febbraio 1914, l'ASCAP (American Society of Composers, Authors and Publishers) è una organizzazione no-profit che protegge i diritti d'autore musicali dei propri membri attraverso un monitoraggio delle esecuzioni pubbliche delle loro composizioni, siano esse via radio o concerti dal vivo.

<sup>3</sup> Il riferimento è alla versione francese del saggio *L'Ouvre d'art à l'époque de sa reproduction mécanisée*, opera di Pierre Klossowski con la decisiva collaborazione dello stesso Benjamin. Si tratta dell'unica pubblicata mentre Benjamin era ancora in vita.

produzione e la riproduzione, poli dialettici di questa singolar tenzone che coinvolge l'arte tutta al tempo della sua condizione di riproducibilità. Passammo dalla perfezione alla perfettibilità dell'opera: proprio la durata ne esce fuori stravolta. Nel corso di questi due ultimi secoli inoltre l'ambiente è cambiato al punto di predisporre al suono come mai in passato: esce allo scoperto per attaccarsi ai corpi, offrendosi tanto al micro dell'individuo quanto al macro della società.

## 2. SEI GRADI

### 2.1 Edison

Se è inutile tentare una scalata alle origini, miriamo a picchettare la storia della musica a partire dal ricorso alla scrittura. Un po' come per le parole, si avvertiva infatti il bisogno di guardarla in faccia, la musica. Ne nacquero sistemi di notazione a partire dalle lettere dell'alfabeto - come quella greca. Le cose cambiarono non poco quando dalle lettere si passò a segni grafici fatti apposta, singolare intreccio di spazio e tempo ad assicurare la ripetibilità della cosa musicale da affidare pur sempre a un esecutore umano.

Così arriviamo al 1877, a celebrare l'anno di nascita del fonografo, sarebbe a dire lo strumento che realizza la grafia musicale al tempo della sua riproducibilità tecnica. Non più disegno del segno per il suono, d'ora in poi disegno del suono direttamente nel segno e senza passare dal via<sup>4</sup>. Disegno del suono programmato dall'uomo per la macchina che lo esegue. E non dimentichiamo che si creano al contempo ed in aggiunta le condizioni per la sua manipolazione; in altre parole si fa un passo in avanti, decisivo verso la *composizione del suono*. Nella storia della musica del '900 il suono si inserisce dunque come fenomeno complesso, evento acustico, sociale, artistico [5].

Torniamo al 1877. Con la sua macchina Edison cambia il ritmo della percezione uditiva: la scrittura e la lettura musicale scorrevano sui due diversi livelli del *qui e ora* da tradurre in *questa volta è per sempre*; come migliore attore non protagonista il pubblico prende finalmente posto, tanto lontano dal musicista quanto più vicino alla sua musica - il qui e ora di artisti e pubblico restava pur sempre lo spazio del concerto, da intendere ancora come quel rituale che limita l'arte alla sua parata sociale [6, 7]. Ultima cosa, non di poco conto: il materiale musicale [8] avrebbe implicitamente guadagnato la totalità dell'ascolto affrancandosi da quella relazione di altezze che reggeva il gioco del suono organizzato da troppi secoli. L'orecchio era finalmente pronto ad accogliere il rumore di fondo della vita quotidiana a rango di materiale musicale: la macchina era stata fortemente didattica in questo senso.

Finalmente si poteva scrivere il suono, direttamente. Era il 17 luglio 1877 quando Edison si rese conto che una puntina, se fatta girare su un disco ad una determi-

nata velocità, emetteva vibrazioni che ricordavano il timbro della voce umana. Una volta esplorata con successo la possibilità di registrare la voce, Edison posiziona senza troppa difficoltà la macchina in ufficio: la vede già amica degli impiegati. Non è un mistero infatti volesse sviluppare il mercato del fonografo attorno all'esercizio della dettatura. Passammo dal dittafono all'industria musicale solo quando si capì l'importanza della musica meccanica elevata a giocattolo alla portata di tutti.

### 2.2 Sousa

John Philip Sousa attraversa sempre in poche righe i libri di storia della musica [9], eppure non mancano ai pari studi specialistici sulla sua vicenda [10, 11] - senza dimenticare la sua autobiografia [12].

Nel testo in questione rileviamo quindi il suo punto di vista interessato alla faccenda della riproducibilità meccanica del suono. Dopo un esordio lirico in bilico tra romanticismo e nostalgia, Sousa ne fa subito una questione di mercato. Avverte che la macchina sia intervenuta irrevocabilmente nel riscrivere i rapporti di fruizione della musica a partire da alcuni momenti specifici - la dichiarazione ad una ragazza, un falò al campeggio, la guerra - che realizzano la vita dell'uomo al netto della sua quotidianità. Il punto è: la macchina allontana l'uomo dal fare musica.

*“La moltitudine di macchine per la riproduzione meccanica, nel loro pazzo desiderio di fornire musica per ogni occasione, è offerta per soppiantare l'illustratore nelle classi, l'orchestra per la danza, i cantanti e gli esecutori pubblici e domestici e così via. Evidentemente si ritiene che nessun campo sia troppo vasto per le loro incursioni, nessuna pretesa troppo stravagante. Ma più possono giudicare queste pretese, più nocivo diventa l'intero sistema”* [2].

In altre parole, è come se l'uomo delegasse la macchina del fare musica, non rivolgendosi più all'uomo. Sousa teme questo. Non solo. Il problema si presenta pure per chi compone: perde definitivamente il controllo della sua musica, bella e fatta per essere ascoltata - senza nemmeno lo sforzo di essere interpretata - dagli altri.

Infine va riconosciuta a Sousa una certa intelligenza nell'orchestrazione. Lavora con le bande, mica con le orchestre. Lavora al tempo della calcificazione del pubblico di massa. Scrive marce militari di tutto rispetto, ancora oggi eseguite. Gente della sua banda - prendete il virtuoso trombonista a coulisse Arthur Pryor - ha reso celebre il glissato esportandolo in Europa tra il 1900 e il 1901. Lo stesso glissato che ritroviamo nel *Pelleas* di Schonberg l'anno dopo. Insomma, era la freschezza di queste sonorità ad imporsi senza destare scalpore nei testi dei critici musicali, trovando spazio semmai nella produzione musicale quotidiana.

E l'America, questa nuova Mecca degli artisti, lavorava proprio in questa direzione: per la musica in conti-

<sup>4</sup> “Phonography as an activity in Edison's formulation comes from the act of inscription involved in writing: “phono-graphy” itself literally means voice writing”. [4]

nuazione<sup>5</sup>. Le radio, insieme alle case discografiche, imponevano i nuovi ritmi di un ascolto dovuto, non più ricercato. Insomma, dopo secoli di musica in tempo reale si imponeva il tempo differito a controbilanciare il peso dei minuti.

A quasi 30 anni dalla realizzazione del fonografo, Sousa interviene dunque per fare il punto della situazione in merito all'intervento del fonografo nello stravolgere del tutto funzione e fruizione della musica.

## 2.2 Benjamin

Walter Benjamin non insiste sulla fruizione e sulla produzione del suono<sup>6</sup>, ma rileva l'urgenza di esporre una teoria dell'arte materialistica a partire dall'immagine in movimento<sup>7</sup>. La stesura del saggio sulla riproducibilità tecnica affianca quell'esposizione materialistica della modernità che Benjamin porta avanti nei *Passages*<sup>8</sup>.

Contemporanea alla stesura del saggio in questione, è l'azione dei totalitarismi a decretare l'urgenza della posizione benjaminina. L'estetizzazione della politica - come sintetizzata nella posizione di Marinetti nel manifesto sulla guerra italo-etiope - propone una intensificazione della tecnica, dei ritmi di vita, dei generatori di energia, tende a un'utilizzazione contro natura delle forze di produzione. Questa tendenza si realizza pienamente nello stato totalitario, laddove *ogni uomo oggi ha il diritto di essere filmato*.

Walter Benjamin fa riferimento alla musica quando deve presentare i confini dell'esponibilità dell'opera; ricorda come sia proprio l'esponibilità a farsi garante nel caso specifico della produzione musicale.

*“E se l'esponibilità di una messa forse non era di per sé minore di quella di una sinfonia, comunque la sinfonia nacque nel preciso momento in cui la sua esponibilità prometteva di diventare maggiore di quella della messa.”*[1]

La riproducibilità tecnica interviene per moltiplicare direttamente le occasioni di esponibilità; nel caso del suono - liberato dalla sua auratica contingenza proprio grazie alla fissazione della sua unicità - il riferimento al solo apparato uditivo avrebbe riscritto l'ambiente insieme al corpo dell'uomo.

L'autore berlinese fa riferimento alla riproduzione tecnica del suono in analogia alla riproduzione delle immagini: laddove il confine tra copia e originale si

disperde, nella musica inizia a farsi spazio il nome dell'interprete.

*“Come nella litografia era virtualmente nascosto il giornale illustrato, così nella fotografia era nascosto il film sonoro. Alla riproduzione tecnica del suono si pose mano alla fine del secolo scorso. Intorno al 1900 la riproduzione tecnica aveva raggiunto uno standard grazie al quale essa non solo iniziò ad assumere come oggetto la totalità delle opere tramandate e a sottoporre il loro effetto alle più profonde trasformazioni, ma si conquistò anche un proprio posto tra i procedimenti artistici.”*[1]

Benjamin pone l'accento sui due tempi della produzione e della riproduzione dell'arte a partire da uno stesso apparato tecnologico: da un lato la memoria del tramandato, dall'altro quello ancora da tramandare possono fronteggiarsi facendo leva sullo standard tecnico raggiunto ad inizio Novecento, quando la percezione viene carsicamente modellata dalle nuove condizioni di fruizione.

## 2.4 Adorno

Adorno si misura con il fenomeno della musica riprodotta nello spirito di una sociologia della musica [16] che rilevi il carattere ancipite di arte e società. Lo fa in alcuni testi dedicati al tema, ai volteggi della puntina sul disco, quando rileva che *c'è da discutere sull'attualità del fonografo* [17]. Siamo in anticipo di pochi anni rispetto al saggio sulla riproducibilità tecnica: tra il 1927 e il 1928 Adorno elabora questi primi appunti, destinati alla pubblicazione solo una quarantina d'anni più tardi.

*“I dischi non sono stati nient'altro che fotografie acustiche riconosciute dal cane in modo tanto gioioso. Non a caso, sia nella fotografia che nella fonografia viene usata l'espressione Platte senz'altra aggiunta. Essa indica il modello di una realtà che si lascia moltiplicare a piacere, si lascia spostare nello spazio e nel tempo e scambiare sul mercato. A causa di ciò essa deve sacrificare la sua terza dimensione: la sua altezza e la sua profondità.”* [17].

Ad Adorno compete una trattazione del fenomeno limitata all'analisi della riproduzione quale luogo in cui il solo mercato può dirsi soddisfatto. La *composizione buona* infatti riflette le mutate condizioni sociali e prova a farsi inascoltabile, garantendo quella spaccatura col pubblico su cui si regge il suo grado di autonomia.

Sull'attualità dei fonografi ci sarebbe da discutere perché, rileva Adorno, il carattere archivistico degli albi discografici è destinato al declino<sup>9</sup>. Considerato ogni fattore atto a radicalizzare la meccanizzazione del suono alla sola riproducibilità, l'autore non avverte l'esigenza di spendere un numero adeguato di parole a so-

<sup>5</sup> Si veda a tal proposito l'analisi proposta dell'industria culturale in *Dialettica dell'Illumismo, quando l'illuminismo diventa mistificazione di massa*. [13]

<sup>6</sup> Tra il 1929 e il 1932 Benjamin, convinto sostenitore di un uso socialmente utile dei nuovi media, lesse alla radio di Berlino e Francoforte una serie di micro-narrazioni rivolte ai ragazzi fra i dieci e i quindici anni. L'esperienza a contatto con medium radiofonico motiva i rapporti di Benjamin con la tecnologia a partire da un uso creativo della stessa, come condizione operativa. [14]

<sup>7</sup> Il materialista storico si accosta ad un oggetto storico solo ed esclusivamente allorché gli si fa incontro come monade. Il film si offriva dunque come monade all'attenzione di Benjamin.

<sup>8</sup> Benjamin si proponeva di risolvere con il *Passagenwerk* il problema centrale del materialismo storico. [15]

<sup>9</sup> L'attuale incremento delle librerie di suoni sembra smentire la posizione di Adorno.

stengo della *soundness* della macchina musicale tale da figurarsi risentito in partenza. Ad esempio, non accetta la forma del disco: in aperta analogia con la fotografia, pesa parecchio la riduzione dell'arte ad oggetto, in un'epoca che riconosce cnicamente il predominio delle cose sull'uomo.

Eppure Adorno non si limita ad attaccare la tecnologia in quanto tale: attacca semmai la supposta neutralità di un mezzo tecnologico in grado di farsi spazio senza troppa difficoltà negli spazi e nei tempi della vita quotidiana, laddove un ometto può fischiettare bonariamente una sinfonia di Beethoven ridotta a motivetto in attesa della metropolitana [18]. L'evento musicale ridotto a merce, ormai pietrificato nella forma del disco, incoraggia una sorta di narcosi collettiva e realizza al massimo grado l'instupidimento dell'udito a contatto con una musica posizionata a misura di ambiente – non più e soltanto a misura d'uomo.

L'influenza di Adorno nel dibattito sulla musica moderna è irrinunciabile, sebbene paghi la riduzione a cliché. Il testo adorniano suggerisce una posizione chiara non solo in riferimento all'analisi formale dell'opera, ma riconosce l'intervento della società quale garante dell'esponibilità dell'opera stessa.

## 2.5 Moholy-Nagy

László Moholy-Nagy collabora con diverse, numerose riviste d'arte. Nel numero 7 della olandese *De Stijl*, anno 1922, presenta *Produktion-Reproduction* [19], un saggio in cui la musica meccanica si era già riscattata dalla sola riproduzione:

*“Poiché la produzione (creatività produttiva) serve soprattutto allo sviluppo dell'uomo, noi dobbiamo cercare di estendere a scopi produttivi quegli apparati (mezzi) finora usati solo a fini a riproduttivi”* [19].

Dove Adorno si limita all'oggetto già lavorato, Moholy-Nagy considera l'oggetto ancora da lavorare: sia l'uomo stesso a praticare scalfiture nel disco di cera

*“senza l'intervento di una azione meccanica esterna, che producano, all'atto della riproduzione, un effetto sonoro, così da rendere possibile, senza nuovi strumenti e senza orchestra un rinnovamento nella produzione sonora (nuovi suoni e rapporti sonori non ancora esistenti) e con ciò contribuire alla trasformazione delle concezioni musicali e delle possibilità compositive”* [19].

Un decennio prima del saggio benjaminiano è Moholy-Nagy dunque a rompere il ghiaccio con l'apparato della macchina che assicura la riproducibilità<sup>10</sup>. L'artista ungherese è mosso dalla rifondazione dell'arte sull'educazione estetica, intesa come educazione all'uso dei sensi, in totale accordo con l'urgenza avvertita da

<sup>10</sup> Una condizione simile è ravvisata da Jean-Luc Godard a proposito dell'invenzione del cinema da parte dei fratelli Lumier [21].

altri artisti<sup>11</sup>. La riproducibilità, non più condizionamento, si fa condizione e diventa luogo di produzione in grado di offrire una risposta adeguata alle sfide percettive poste dalla modernità.

Non sorprenda la sua proposta per un uso creativo - non più ricorsivo - del disco. Moholy-Nagy insegna al Bauhaus: la scuola è il luogo dell'artista il cui compito sociale è l'insegnamento. Ne viene fuori un costante ottimismo nei confronti della possibilità di un uso estetico e umanistico della tecnologia [20]: senza ancora aver letto Benjamin, Moholy-Nagy<sup>12</sup> anticipa il tema del gioco, così come trova spazio nel saggio sulla riproducibilità quale condizione operativa della seconda tecnica<sup>13</sup>. Parla infatti della possibilità di usare il grammofo per produrre suoni *ex novo*, senza alcuno strumento, anziché registrarli e renderli riascoltabili come voleva l'uso tradizionale.

Il riferimento al suono, al caso del grammofo appare in nota, laddove il corpo del testo si concentra sulla produzione fotografica: non è un mistero che lo stesso Moholy-Nagy sia stato tra i primi a ricorrere al fotomontaggio per combinare diverse immagini fotografiche - avulse dal contesto, ritenute insignificanti o banali, venivano ricomposte in un nuovo contesto, significativo.

Il caso di Moholy-Nagy anticipa la relazione tra visivo e sonoro in rapporto alla meccanizzazione dell'opera d'arte, tema ricorsivo dell'estetica contemporanea. Il suo studio su *Pittura Fotografia Film* nasconde il riferimento al movimento, a quella percezione in movimento che evidenzia la natura del fenomeno sonoro, costretto ad estendersi non solo nell'ambiente. La traduzione del tempo in spazio ad opera delle tecnologie infittisce la relazione tra suono e immagine a partire dalla leva comune della riproducibilità, nella tensione che si realizza tra produzione e riproduzione, laddove non c'è spazio per la copia se non come identità dell'originale.

## 2.6 McLuhan

McLuhan radicalizza l'indicazione benjaminiana sulla tecnologia come apparato e finisce col posizionare il medium come estensione dell'uomo: la percezione trova spazio sul corpo stesso dell'individuo, non è più limitata ai soli sensi ma viene incoraggiata dall'ambiente che lo circonda. Lo capiamo bene nel capitolo dedicato al grammofo, sottotitolato dal me-

<sup>11</sup> Gli fa eco Varèse, che nel corso di una tavola rotonda dal titolo *la Meccanizzazione della musica* del 1930 riferisce: “Non è detto che con l'educazione un orecchio poco sviluppato all'inizio non possa migliorare le proprie facoltà uditive” [22].

<sup>12</sup> Gli autori sono accumulati da un uso socialmente utile dei nuovi media.

<sup>13</sup> “Una volta per tutte – questa è stata la parola d'ordine della prima tecnica. Una volta non conta – questa è la parola d'ordine della seconda tecnica. L'origine della seconda tecnica deve essere ricercata nel momento in cui l'uomo, guidato da un'astuzia inconscia si apprestò per la prima volta a prendere le distanze dalla natura. In altri termini: la seconda tecnica nacque come gioco” [23].

diologo canadese *Il giocattolo che ha contratto il petto della nazione*<sup>14</sup>.

Nel presentare l'oggetto McLuhan rende il lettore partecipe dell'accoglienza tributata da John Philip Sousa al tempo della sua comparsa sul mercato americano:

*“Il fonografo farà passare di moda gli esercizi vocali! Che ne sarà della gola della nazione? Non si indebolirà? Che ne sarà del petto della nazione? Non si contrarrà?”* [3].

Il nuovo mezzo nasceva dalla voce per farsene estensione: poteva tranquillamente ridurre l'attività vocale degli individui, l'educazione musicale veniva così tradita a vantaggio del solo ascolto. E soprattutto il grammofo, in virtù del suo calore al pari della radio, poteva lentamente scalfire l'inconscio acustico grazie ad una trasgressiva invisibilità.

La storia del grammofo offre infatti, secondo McLuhan, la riduzione della distanza tra forma e funzione, condizione generativa dei media elettrici per via del loro carattere di campo totale:

*“La breve storia del grammofo comprende tutte le fasi della parola scritta, stampata e meccanicizzata”* [3].

McLuhan rileva la necessità di considerare il fenomeno sonoro al netto della sua predisposizione ambientale e sociale, convinto dell'intreccio tra musica e telefonia, musica e radiofonia, musica in quanto onde elettromagnetiche per una nuova misurazione del mondo dei suoni. In effetti, lui stesso ricorda come

*“Edison aveva pensato al fonografo come un ripetitore del telefono. [...] L'apparecchio per suonare i dischi era immaginato come una specie di registratore delle conversazioni telefoniche”*<sup>15</sup> [3].

Insomma, la macchina è stata piegata ad un utilizzo musicale solo quando fu considerata come mezzo di svago. L'incapacità di prevedere questa funzione del fonografo era di fatto un'incapacità di intendere il significato della rivoluzione elettrica in genere.

*“Oggi accettiamo pacificamente il grammofo come un giocattolo e un sollievo, e hanno acquistato lo stesso carattere di svago anche il giornale, la radio e la tv. Nello stesso tempo lo svago portato all'estremo diventa la forma principale degli affari e della politica.”* [3].

McLuhan riparte proprio da Edison, fa rapporto al Finnegans Wake di Joyce<sup>16</sup> e chiude con una considerazione che anticipa gli spazi e i tempi della multimedialità.

La traduzione della scrittura in suono compie un passo avanti nello spazio auditivo prima, in quello visivo poi: non ci sono più mura a misurare lo spazio - la corrente elettrica le ha abbattute pur lasciandole sul posto.

L'oggetto scompare al cospetto dell'informazione sotto forma di energia elettrica.

### 3. CONCLUSIONE

In merito ai rapporti tra la macchina e il suono è stato possibile tracciare un resoconto a partire da *La minaccia della musica meccanica*, il saggio di John Philip Sousa del 1906 dalla cui lettura è seguito il contributo proposto.

Lontani più di un secolo dalla sua stesura, è opportuno segnalare come possa ancora offrire un riferimento storico adeguatamente documentato sulla questione della musica a misura di macchina. Portato all'attenzione della comunità elettroacustica nel 1993 da Curtis Road [24], *The menace of Mechanical Music* dialoga a più riprese con la vita quotidiana, a segnalare il carattere fortemente interattivo tra macchina e uomo già nel processo di fruizione - non solo di composizione - del suono, della musica, del paesaggio sonoro.

Gli artisti e teorici citati insieme al compositore americano hanno studiato l'impatto culturale dei media marcando il nesso tra storia dei media e storia della percezione, tra storia degli apparecchi tecnici e storia delle forme dell'esperienza sensibile - un'esperienza che, come scrive Benjamin nel saggio sulla riproducibilità tecnica, si svolge all'interno del «medium della percezione». Tutti gli autori chiamati in causa hanno dunque tradotto singolarmente le regole della riproducibilità tecnica applicate ad una traccia audio, laddove il tema della reversibilità e della ripetibilità della stessa sembrerebbe riverberarsi sulla memoria del fruitore. In questo modo un certo problema quantitativo ha iniziato ad emergere tacitamente, trasformando il primato dell'esperienza auratica nella definizione di una coazione a ripetere quale leva della produzione musicale che nasce col solo scopo di essere riprodotta.

In conclusione, il compositore e direttore di banda statunitense ha scritto di macchine parlanti molto prima di Alexa<sup>17</sup>, si è preoccupato della qualità della produzione musicale in modo completamente diverso che le avanguardie, ha avuto paura di cosa ne sarebbe stato della guerra, dell'innamoramento nell'epoca della musica meccanica, si è chiesto inoltre cosa ne sarebbe stato della didattica della musica per arrivare alle conclusive preoccupazioni espresse a riguardo del diritto d'autore - specifica traduzione nel campo dell'arte del sempre più scottante tema della sicurezza.

<sup>14</sup> Il riferimento all'impiego nello svago diventa evidente già dal sottotitolo che riprende il saggio di Sousa *The menace of mechanical music*.

<sup>15</sup> Un esempio è offerto dalle note vocali della messaggistica istantanea.

<sup>16</sup> Non avrebbe potuto ideare quest'ore in un'epoca che non fosse quella che aveva prodotto il grammofo e la radio [3].

<sup>17</sup> Si fa riferimento all'avvento e allo sviluppo sempre più massiccio delle tecnologie di sintesi vocale e delle interfacce conversazionali integrate nei dispositivi elettronici domestici - gli assistenti digitali Amazon Alexa, Google Home - e mobili - Siri di Apple o Cortana di Microsoft - o nelle autovetture - l'assistente di guida Dragon Drive di Toyota [25].

Insomma, si è preoccupato della musica inserita meccanicamente nel ritmo della quotidianità al pari di compositori presenti 112 anni dopo.

#### 4. REFERENCES

- [1] W. Benjamin, *Aura e choc*, Einaudi, Torino, 2012.
- [2] J.P. Sousa, *The menace of mechanical music*, originally published in *Appleton's Magazine*, Vol. 8, 1906, pp. 278-284, [www.phonozoic.net/n0155.htm](http://www.phonozoic.net/n0155.htm).
- [3] M. McLuhan, *Gli strumenti del comunicare*, Il saggiautore, Milano, 1967.
- [4] D. Kahn, "The Sound of Music", *Der freie Klang symposium, Ars Electronica Festival*, 1987.
- [5] Aa. Vv., *Nuova Atlantide: il continente della musica elettronica*, La Biennale di Venezia, Venezia, 1986.
- [6] A. Vidolin, *Suonare lo spazio elettroacustico* (articolo on-line. Materiale didattico per il corso di *Sistemi di Elaborazione per la Musica*, Dipartimento di elettronica e informative, facoltà di Ingegneria, Università degli studi di Padova; Prof. G. De Poli), 2001, [pagg.9\(www.dei.unipd.it/~musica/Dispense/VidolinSuonare.pdf\)](http://www.dei.unipd.it/~musica/Dispense/VidolinSuonare.pdf).
- [7] F. Delalande, *Il paradigma elettroacustico*, in *Enciclopedia della musica*, Vol. I, Il Novecento, Einaudi, Torino, 2001.
- [8] A. Di Scipio, *Teoria e prassi della musica elettronica*, La Terza, Bari, 1995.
- [9] A. Ross, *Il Resto è rumore. Ascoltando il XX secolo*, Bompiani, Milano, 2011, p. 145.
- [10] P. E. Bierly, *John Philip Sousa: american phenomenon*, Appleton-Century-Crofts, New York, 1973.
- [11] P. E. Bierly, *The incredible band of John Philip Sousa*, University of Illinois Press, Urbana, 2006.
- [12] J. P. Sousa, *Marching along, recollections of men, women and music*, a cura di P. E. Birley, Integrity Press, Westerville, 1994.
- [13] T. W. Adorno, M. Horkeimer, *Dialettica dell'illuminismo*, Einaudi, Torino, 2010.
- [14] W. Benjamin, *Burattini, streghe e altri briganti. Racconti radiofonici per ragazzi*, BUR, Milano, 2014.
- [15] W. Benjamin, *I Passages di Parigi*, Einaudi, Torino, 2000.
- [16] T.W. Adorno, *Introduzione alla sociologia della musica*, Einaudi, Torino, 1971.
- [17] T.W. Adorno, *Long Play e altri volteggi della puntina*, Castelvechi, Roma, 2012.
- [18] T.W. Adorno, *Sulla popular music*, Castelvechi, Roma, 2008.
- [19] L. Moholy-Nagy, *Pittura Fotografia Film*, Einaudi, Torino, 2010.
- [20] A. Mastrogiacomo, *Senza mura. Sui confini della multimedialità*, Fogli e Parole d'arte, Roma, 2018 (<https://www.foglidarte.it/testuali-parole/680-senza-mura-sui-confini-della-multimedialita.html>).
- [21] J-L. Godard, *Introduzione alla vera storia del cinema*, PGrego, Milano, 2012.
- [22] E. Varèse, *Il suono organizzato*, Lim, Lucca, 1989.
- [23] W. Benjamin, *L'opera d'arte nell'epoca della sua riproducibilità tecnica*, Donzelli, Roma, 2012.
- [24] C. Roads, "Introduction to 'The Menace of Mechanical Music' by John Philip Sousa", *Computer Music Journal*, Vol. 17, No. 1 (Spring, 1993), The MIT press, Cambridge, pp. 12-13.
- [25] D. Napolitano, "Voci digitali per desideri reali: suono e soggettività nell'epoca delle macchine parlanti", in *d.a.t. [divulgazione audiotestuale]*, n. 2, anno II, aprile 2018, Napoli, pp. 87-106.

# QUALE FUTURO PER IL FORMATO IEEE 1599?

F. Avanzini<sup>1</sup>, A. Baratè<sup>1</sup>, G. Haus<sup>1</sup>, L.A. Ludovico<sup>1</sup>, D.A. Mauro<sup>2</sup>, S. Ntalampiras<sup>1</sup>, G. Presti<sup>1</sup>

<sup>1</sup> LIM – Laboratorio di Informatica Musicale, Dipartimento di Informatica, Università degli Studi di Milano  
Via Celoria, 18 – 20133 Milano, Italia

lim@di.unimi.it

<sup>2</sup> Department of Computer and Information Technology, Marshall University  
1548 6th Ave. Apt. 19 – Huntington, WV, 25701 (USA)

maurod@marshall.edu

## ABSTRACT

Il formato IEEE 1599 è uno standard internazionale per la codifica multistrato dell'informazione musicale. Il LIM – Laboratorio di Informatica Musicale dell'Università degli Studi di Milano è il gruppo di lavoro che maggiormente ha contribuito alla promulgazione del formato e allo sviluppo di un ecosistema di applicativi per la generazione, manipolazione e fruizione dell'informazione codificata secondo tale standard. A dieci anni dal rilascio ufficiale delle specifiche dell'IEEE 1599, nel presente lavoro ci si interroga sui punti critici che ne hanno limitato la diffusione e sulle nuove direzioni da intraprendere per far evolvere il formato e contribuire al suo futuro successo.

## 1. INTRODUZIONE

IEEE 1599 è un linguaggio basato su XML che mira a fornire una descrizione quanto più esaustiva di un singolo brano musicale. Tale formato costituisce un ambiente di rappresentazione *multistrato* che articola l'informazione su più *strati* o *livelli*: il livello generale (*general*) si occupa principalmente dei metadati catalografici, il livello logico (*logic*) della rappresentazione astratta della partitura in termini di simboli musicali, il livello strutturale (*structural*) della descrizione delle relazioni tra oggetti musicali, il livello notazionale (*notational*) delle istanze grafiche associabili a una partitura, il livello di performance (*performance*) dei formati per la riproduzione sintetica della musica, e infine il livello audio (*audio*) delle registrazioni delle esecuzioni.

IEEE 1599 integra la rappresentazione dell'informazione in XML con standard e formati di file già definiti e comunemente accettati, tipicamente applicati a singoli livelli: i formati grafici per gli aspetti notazionali, i formati audio per le registrazioni, e via dicendo. Lo scopo è recuperare il vasto patrimonio documentale già disponibile e integrarlo all'interno di una codifica unica. Da questo punto di vista, IEEE 1599 si pone sia come formato di interscambio tra

diverse applicazioni, sia come aggregatore per contenuti legati allo stesso brano musicale.

Il formato IEEE 1599, però, non è un semplice raccoglitore di informazioni eterogenee. Uno dei suoi obiettivi primari è far emergere, ove possibile, relazioni spazio-temporali tra differenti rappresentazioni degli eventi musicali descritti. Innanzi tutto, chiariamo il concetto di *evento musicale*, lasciato volutamente vago nelle specifiche di IEEE 1599: si tratta di un'entità informativa, potenzialmente di qualsiasi livello di astrazione e/o aggregazione, considerata significativa da parte dell'autore della codifica ai fini della specifica situazione. A titolo di esempio, risulta scelta comune prendere in considerazione tutti i simboli di accordo (eventualmente degenerare) e di pausa contenuti in una partitura secondo la notazione tradizionale. In realtà, è facoltà dell'autore della codifica effettuare scelte radicalmente diverse; questo perché la partitura potrebbe non avere una notazione intesa in senso tradizionale (si pensi al repertorio della musica elettronica), oppure potrebbe risultare troppo onerosa una codifica con granularità alla singola nota (si pensi a un brano di grande estensione), o semplicemente potrebbe essere altro l'interesse musicale o musicologico (si pensi a un'analisi armonica, in cui gli eventi musicali di rilievo sono le situazioni accordali).

Una volta effettuata una scelta in tal senso, gli eventi musicali vengono marcati univocamente attraverso l'associazione di un codice identificativo. La struttura dati delegata a tale scopo è detta *spine*, dato che costituisce la vera e propria spina dorsale della codifica. Difatti, gli identificativi serviranno poi da riferimento per i rimandi presenti nei vari livelli, che ospitano descrizioni eterogenee per i singoli eventi musicali [1].

A scopo di esempio, si consideri un evento musicale marcato dall'identificativo  $e_n$ . Tale etichetta sarà presente nello *spine*, in modo da consentire ai vari livelli di fare gli opportuni riferimenti a  $e_n$ . All'interno del livello notazionale, sarà possibile riferire una data area della partitura alla descrizione grafica di  $e_n$ , mentre nello strato audio si potrà denotare un dato intervallo temporale come l'occorrenza sonora di  $e_n$ . Il principio appena enunciato crea la cosiddetta *sincronizzazione inter-livello* interna a un documento IEEE 1599, in quanto i molteplici rimandi – a partire da strati differenti – allo stesso identificativo dello *spine* di fatto generano una relazione spazio-temporale tra le descrizioni logiche, grafiche, audio, strutturali dello stesso

Copyright: ©2018 F. Avanzini<sup>1</sup>, A. Baratè<sup>1</sup>, G. Haus<sup>1</sup>, L.A. Ludovico<sup>1</sup>, D.A. Mauro<sup>2</sup>, S. Ntalampiras<sup>1</sup>, G. Presti<sup>1</sup> et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution License 3.0 Unported](https://creativecommons.org/licenses/by/3.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

evento musicale.

A questo meccanismo si sovrappone poi la *sincronizzazione intra-livello*, che ha luogo quando un unico strato ospita più istanze della stessa tipologia di materiale multimediale. Si pensi alla disponibilità di più tracce audio o più versioni della partitura per la stessa composizione. Questi materiali saranno rappresentati come istanze differenti all'intero dello stesso livello, e ciascuna di queste ultime conterrà presumibilmente una descrizione differente per  $e_n$ . Per approfondimenti su approcci e tecniche di sincronizzazione in IEEE 1599, si veda il lavoro [2].

L'unione delle varie caratteristiche di IEEE 1599 – scelta del livello di astrazione della descrizione, strutturazione gerarchica di informazione proveniente da molteplici fonti, sincronizzazione inter-livello e intra-livello – comporta un cambio di paradigma nella rappresentazione digitale dell'informazione musicale. Per quanto riguarda i possibili risvolti pratici, lo standard si pone come punto di riferimento per qualsiasi applicazione che possa trarre beneficio da una visione multisfaccettata di un brano musicale: dai *media player* con funzioni avanzate di *score following* ai sistemi per l'apprendimento della musica basati sull'uso del computer, dalla valorizzazione di archivi musicali tradizionali attraverso le nuove tecnologie alla composizione assistita e alla musicologia computazionale.

Dunque, attraverso lo sviluppo di opportune applicazioni, il formato IEEE 1599 potrebbe spostare le frontiere dell'*entertainment*, della formazione, dell'editoria e della ricerca musicale. Eppure, nei 10 anni dal rilascio dello standard, il formato non ha conosciuto la diffusione che sarebbe stato lecito attendersi. Lo scopo del presente lavoro è indagare le cause di questo fenomeno e proporre nuove direzioni da intraprendere per far evolvere il formato e contribuire al suo futuro successo.

Il presente lavoro si articola nel seguente modo: la Sezione 2 evidenzia le tappe fondamentali nello sviluppo di IEEE 1599, la Sezione 3 fa un'analisi critica delle problematiche nell'adozione dello standard, la Sezione 4 descrive lo stato dell'arte, e infine la Sezione 5 ne preconizza le prospettive a breve e medio termine.

## 2. CRONISTORIA DI IEEE 1599

La storia del formato IEEE 1599 ha inizio nel 1992, con la costituzione dell'*IEEE Computer Society Task Force on Computer Generated Music*, evolutosi poi nel 1994 nello *IEEE Technical Committee on Computer Generated Music (TCCGM)*. L'IEEE, acronimo di *Institute of Electrical and Electronic Engineers*, è un'associazione internazionale di scienziati e professionisti che ha come obiettivo la promozione delle scienze tecnologiche. Tale organizzazione si occupa, tra le altre cose, di promulgare standard industriali riconosciuti a livello internazionale; per questo motivo, si è pensato di costituire all'interno dell'IEEE un gruppo di lavoro sulla rappresentazione della musica in ambito digitale. I fondatori del TCCGM sono stati Denis Baggi e Goffredo Haus, cui va la paternità dell'idea di strutturare l'informazione musicale secondo un approccio multistrato.

Un anno cardine nello sviluppo di IEEE 1599 è stato il 2001, con la creazione dell'*IEEE Standards Association*

*Working Group on Music Application of XML* che ha individuato proprio nel meta-linguaggio XML il formato più idoneo per la codifica, grazie ai molteplici vantaggi offerti. Sempre nel 2001 ha avuto luogo l'approvazione della *Recommended practice for the Definition of a Commonly Acceptable Musical Application Using the XML Language* da parte dell'*IEEE Standard Association*, primo passo formale sulla strada della standardizzazione.

Nel 2002 la comunità scientifica è stata chiamata a raccolta presso l'Università degli Studi di Milano per la prima edizione dell'IEEE International Conference on Musical Application using XML (MAX 2002) [3]. In questa occasione è stata presentata una prima bozza del formato, con il nome provvisorio di MX.

Nel 2008 si è conclusa la fase di *balloting*, ossia la valutazione finale del nuovo formato da parte di una platea internazionale di circa 100 esperti del settore. Il risultato è stato ampiamente positivo. Nasceva dunque lo standard IEEE 1599, evoluzione del precedente formato MX.

Il 2011 ha segnato una tappa importante nello sviluppo di applicativi basati su IEEE 1599, grazie al finanziamento del progetto *Enhanced Music Interactive Platform for Internet User (EMIPU)* che ha portato alla realizzazione di un portale web tuttora disponibile e, soprattutto, al rilascio del primo *web viewer* completamente funzionante.

Nel 2013 è stato pubblicato dalla casa editrice Wiley un volume dedicato a IEEE 1599 [4]. Nel frattempo, si sono moltiplicate le pubblicazioni scientifiche sull'argomento e le manifestazioni (quali mostre, conferenze, ecc.) in cui le potenzialità del formato sono state svelate al grande pubblico.

Allo stato attuale si contano circa 50 lavori presentati a convegno o pubblicati su rivista incentrati sul formato.<sup>1</sup>

Nel 2018, in occasione del decennale dal rilascio ufficiale dello standard, l'IEEE ha chiesto di ricostituire il working group al fine di aggiornarlo sulla base dell'esperienza accumulata e delle nuove sfide tecnologiche. Di questo argomento si parlerà più diffusamente nella Sezione 5.

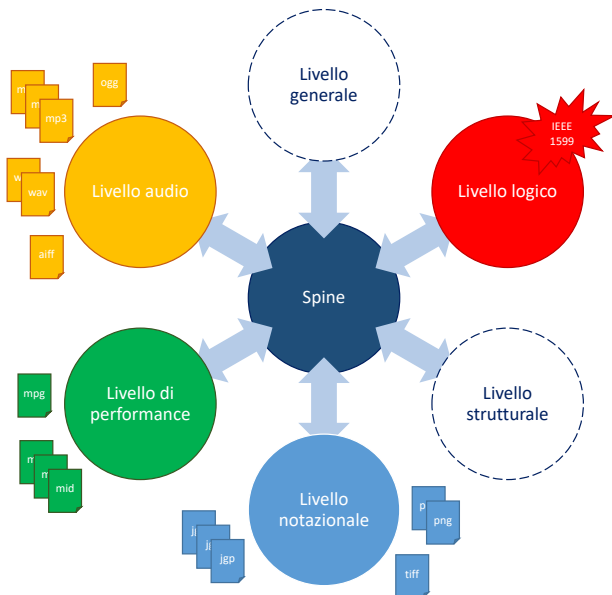
## 3. CRITICITÀ DI IEEE 1599

I principali aspetti di criticità del formato IEEE 1599 sono stati analizzati e condivisi con la comunità scientifica [5]. Si tratta di elementi emersi dopo la fase di standardizzazione, quando il rilascio di applicazioni e un confronto più serrato con il mondo dell'accademia e dell'industria ha evidenziato le prime lacune.

### 3.1 Più formati per il livello logico

Un primo aspetto di criticità è una forma di asimmetria che emerge nell'approccio di IEEE 1599, in particolare per quanto riguarda la descrizione logica dei simboli musicali. Una delle intuizioni del formato è consistita nell'accettare file esterni per gli oggetti multimediali collegati, quali immagini, audio e video, supportando i principali formati in uso. Questa soluzione presenta diversi vantaggi, tra cui

<sup>1</sup> L'elenco completo, costantemente aggiornato, si trova all'indirizzo [http://iee1599.lim.di.unimi.it/documentation\\_papers.php](http://iee1599.lim.di.unimi.it/documentation_papers.php).



**Figura 1.** Rappresentazione grafica dei livelli IEEE 1599. Tralasciando gli strati generale e strutturale, che usualmente non prevedono l’aggancio di contributi esterni, i livelli notazionale, di performance e audio supportano il collegamento di molteplici oggetti digitali, ciascuno con il proprio formato di codifica, mentre il livello logico consente di descrivere l’informazione musicale solo nel formato XML proprietario, generando dunque una situazione asimmetrica.

la separazione tra elementi descrittivi strutturali (espressi come sintassi XML) e istanze multimediali (file esterni), il recupero di un vasto *corpus* documentale di oggetti digitali già codificati in formati comunemente accettati, e la conseguente possibilità di non utilizzare la sintassi XML – notoriamente verbosa – per descriverli, appesantendo ulteriormente la codifica XML. Sfortunatamente, al contrario di altri strati aperti a una molteplicità di tecniche di codifica, il livello logico supporta solo una forma di rappresentazione proprietaria. Questo impedisce di collegare file esterni codificati secondo standard comunemente accettati in questo ambito: limitandosi alle codifiche XML, basti pensare a MusicXML [6] e al formato MEI [7]. La Figura 1 fornisce una rappresentazione grafica della situazione, prendendo ad esempio un documento particolarmente ricco di collegamenti con oggetti multimediali esterni.

A causa di questa limitazione, non si possono recuperare in maniera diretta i vasti archivi di partiture già disponibili in tali formati (dovendo passare attraverso la creazione di moduli software di conversione), né si possono utilizzare software di editing della partitura quali Finale, MuseScore e Sibelius che producono nativamente esportazioni in MusicXML (dovendo rilasciare plugin di esportazione appositamente scritti); aspetto non secondario, il gruppo di lavoro IEEE 1599 non può beneficiare, se non per via indiretta, della sistematica opera di analisi critica e di affinamento dei formati citati da parte delle rispettive comunità.

### 3.2 Strutturazione e caratterizzazione delle istanze

Un secondo punto critico, non adeguatamente analizzato in fase di progettazione del formato, riguarda il concetto di istanza multipla all’interno di uno strato. Il livello notazionale, ad esempio, presenta una gerarchia a più stadi per organizzare i file grafici: un primo stadio è quello della raccolta, utile a racchiudere tutte le scansioni che si riferiscono a una stessa edizione; un secondo stadio è il singolo file. Questa soluzione potrebbe sembrare idonea, ma viene messa in crisi ad esempio dalla disponibilità della partitura completa di una certa edizione e delle relative parti orchestrali. Questa situazione implica che il livello notazionale racchiuda  $n$  raccolte di file grafici, aventi tutte la stessa relazione gerarchica, di cui un sottoinsieme  $m$  è potenzialmente legato ad uno specifico raggruppamento, il che evidentemente non emerge dalla struttura.

La questione si fa, se possibile, ancora più critica per il livello audio, in cui non sono previsti sottolivelli di raggruppamento. Dunque, delle  $k$  istanze audio ivi contenute, alcune potrebbero riferirsi a esecuzioni da parte di interpreti differenti, altre potrebbero essere esecuzioni in qualche modo collegate tra di loro (ad esempio, registrazioni dello stesso cast/organico in serate differenti), altre ancora potrebbero avere un rapporto gerarchico rispetto ad una specifica istanza (ad esempio, registrazioni multi-traccia rispetto al master stereo).

### 3.3 Gestione della proprietà intellettuale

Un ulteriore aspetto di criticità sta nella scarsa attenzione posta, ai tempi della stesura delle specifiche, alla tematica del *digital rights management* (DRM), particolarmente rilevante per un formato che fa dell’aggregazione di contenuti multimediali potenzialmente protetti il suo punto di forza. Da un lato, è pur vero che il formato è stato sviluppato in un’epoca in cui non si era ancora affermato uno standard prevalente per la rappresentazione in XML delle licenze, mentre al giorno d’oggi esistono formati quali MPEG-21 REL e ODRL. Inoltre, va sottolineato che il tipo di fruizione supportata da IEEE 1599 funge da protezione implicita rispetto all’esperienza utente: pur essendo tecnicamente possibile scaricare in locale i singoli file grafici e registrare i flussi multimediali, non è comunque consentito ottenere le informazioni di sincronizzazione, vero valore aggiunto del formato, né riprodurre la sessione interattiva esperita dall’utente [8].

In ogni caso, i detentori dei singoli oggetti digitali, ossia *content provider* e *stakeholder*, mostrano una comprensibile preoccupazione verso i propri beni da tutelare, e la pressoché totale assenza di meccanismi di protezione rende il formato poco attraente ai loro occhi.

Va inoltre sottolineato che i detentori dei diritti da coinvolgere nella creazione di un documento IEEE 1599 complesso sono spesso figure giuridicamente distinte. A titolo esemplificativo, si consideri un archivio pubblico che conserva il manoscritto originale, una o più case editrici musicali che hanno in catalogo diverse partiture a stampa, etichette discografiche detentrici dei diritti sull’audio, forme contrattuali eterogenee pattuite con gli artisti, e via dicendo.

Peraltro, allo stato attuale IEEE 1599 non prevede meccanismi sintattici per associare codici meccanografici o commerciali – quali ISRC, ISBN o EAN – agli oggetti che ne dispongono.

### 3.4 Altre criticità

A questi punti cruciali si aggiungono alcuni aspetti secondari, ma da prendere comunque in considerazione nell’ottica di una maggiore usabilità e diffusione di IEEE 1599. Ad esempio, si avverte l’assenza di un formato di file standardizzato che funga da contenitore per i diversi oggetti digitali connessi alla descrizione di un singolo brano. Come si è detto, al documento principale in XML si affiancano codifiche in formato grafico, audio e video, e sarebbe opportuno poter distribuire un singolo pacchetto anziché molteplici file.

Allargando lo sguardo a un orizzonte più ampio, essendo il singolo documento focalizzato sulla descrizione di un brano musicale, si avverte l’assenza di un meccanismo sintattico per organizzare insieme di codifiche IEEE 1599 (ad esempio, la collezione relativa a un’intera opera lirica o ai movimenti di una sinfonia).

### 3.5 Tempi di preparazione dei materiali

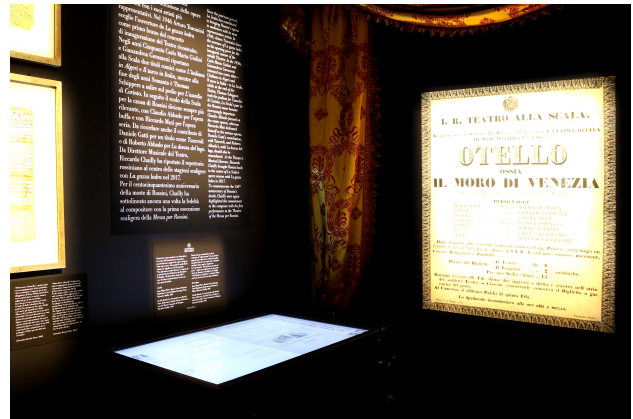
Infine, va citato un aspetto critico che mina l’appetibilità di IEEE 1599 da parte dell’industria. Non si tratta di una problematica strettamente collegata alle caratteristiche del formato, quanto piuttosto all’assenza di strumenti automatici per l’allineamento dell’informazione logica con le rappresentazioni grafiche e audio del brano.

Una stima puntuale dei tempi di lavorazione richiesti per la produzione di un documento IEEE 1599 in generale non è possibile, a causa dell’elevato numero di fattori che la potrebbero influenzare in maniera significativa: la partitura (livello logico) è già disponibile in un qualche formato simbolico? Quanto è estesa temporalmente e quanto è complessa dal punto di vista notazionale? Quanti oggetti multimediali si vogliono agganciare alla descrizione, e con quale livello di granularità?

Volendo comunque fornire una stima approssimativa basata sull’esperienza, allo stato attuale i tempi di lavorazione per una composizione di media complessità e di estensione limitata (indicativamente da 3 a 5 pagine) cui agganciare 3 versioni della partitura (livello notazionale) e 3 tracce (livello audio) sono circa 3 giorni/uomo, accettabili per la realizzazione di un dimostrativo ma probabilmente eccessivi per il rilascio di un’opera più articolata. Mettendo a punto algoritmi per il riconoscimento automatico, i tempi di lavorazione subirebbero un drastico abbattimento, rendendo l’operazione più sostenibile anche dal punto di vista economico.

## 4. IL PRESENTE DI IEEE 1599

Allo stato attuale, il formato IEEE 1599 trova applicazione in un insieme significativo di campi che include la fruizione multimediale [9], l’educazione musicale [10] e la valorizzazione dei beni culturali immateriali [11].



**Figura 2.** Dimostrativo installato presso il Museo teatrale alla Scala basato sul formato IEEE 1599.

Ad esempio, lo standard viene regolarmente utilizzato per la realizzazione di dimostrativi all’interno di mostre ed eventi pubblici. Da questo punto di vista, vale la pena citare la collaborazione di lungo termine con istituzioni quali il Ministero per i Beni e le Attività Culturali, l’Archivio Storico Ricordi, il Teatro alla Scala di Milano, e – in tempi più recenti – Palazzo Reale di Milano e Paul Sacher Stiftung.<sup>2</sup> In Figura 2 viene mostrato l’applicativo attualmente installato presso il Teatro alla Scala di Milano nell’ambito della mostra “Gioachino Rossini al Teatro alla Scala”.

Basate su IEEE 1599 sono anche le applicazioni avanzate per la fruizione e l’interazione con i contenuti musicali messe a punto per Bach-Archiv Leipzig, Orchestra Verdi di Milano, Pearson, RAI Radiotelevisione Italiana, RSI Radiotelevisione svizzera di lingua italiana.

Nei 10 anni trascorsi dalla standardizzazione del formato, sono stati realizzati diversi motori di sincronizzazione e riproduzione del formato IEEE 1599. Al fine di valorizzare l’esperienza utente supportata dalle caratteristiche del formato, tali software propongono tre modalità di navigazione dei contenuti musicali:

1. *score following* – Agendo sugli usuali controlli propri di qualsiasi media player (Play, Pause, ecc.), l’u-

<sup>2</sup> Un elenco delle principali manifestazioni si trova all’indirizzo [http://iee1599.lim.di.unimi.it/practice\\_exhibitions.php](http://iee1599.lim.di.unimi.it/practice_exhibitions.php).

tente ha la possibilità di ascoltare una traccia audio e visualizzare una partitura a propria scelta tra quelle collegate al brano, seguendo il punto corrente di esecuzione attraverso la visualizzazione di un opportuno cursore che sincronizza la parte grafica con l'audio. Pur limitandosi a funzionalità basilari, questa prima modalità già introduce alcune caratteristiche avanzate, tra cui la disponibilità di molteplici oggetti digitali grafici e audio tra cui scegliere, o la flessibilità del cursore, non una linea verticale che scorre sull'accollatura ma piuttosto un insieme di aree dalla forma configurabile per tener conto di eventuali disallineamenti nella scrittura o di notazioni non tradizionali;

2. *interazione con i contenuti* – Oltre ai controlli standard per modificare l'istante di riproduzione (l'azione sulla barra di avanzamento dell'audio, ad esempio) o per la voltata di pagina, si sono rese sensibili al click tutte le aree della partitura contenenti simboli musicali, rendendo dunque molto intuitivo il riposizionamento, con immediata risincronizzazione dei materiali, su un punto desiderato;
3. *sostituzione in tempo reale dei contenuti* – Avendo a disposizione una codifica ricca di oggetti digitali collegati, è possibile sostituire in tempo reale i contenuti in uso mantenendo l'istante di esecuzione corrente. Questo permette di confrontare in modo istantaneo diverse performance o differenti versioni della partitura.

In tempi recenti il team di sviluppo si è dedicato in particolare alla fruizione via web dei documenti IEEE 1599, che pone ulteriori sfide tecnologiche, quali ad esempio l'invio di molteplici flussi multimediali al client [12]. Un primo prototipo è costituito dalla piattaforma EMPIU, sviluppata e rilasciata nel 2012 grazie a un co-finanziamento di Regione Lombardia e regione francese di Languedoc-Roussillon.<sup>3</sup>

Da pochi mesi è disponibile una versione aggiornata del portale, dedicata alla documentazione del progetto IEEE 1599 e alla visualizzazione web di un sottoinsieme significativo di esempi, sia musicali sia extra-musicali. Si tratta tuttora di un *work in progress* che però introduce caratteristiche interessanti, quali la possibilità di configurare lo spazio di lavoro ed affiancare o sovrapporre un numero arbitrario di finestre contenenti oggetti multimediali sincronizzati. In tal modo, è possibile ad esempio seguire la notazione musicale contemporaneamente su diverse versioni della partitura. Inoltre, la piattaforma presenta alcuni strumenti di analisi musicologica e di visualizzazione alternativa rispetto alla notazione tradizionale, il tutto all'interno di un ambiente di navigazione multistrato e completamente sincronizzato. Il nuovo *viewer* è disponibile all'indirizzo <http://ieee1599.lim.di.unimi.it/>, nell'area Music archive. La sua interfaccia è mostrata in Figura 3.

Se, lato utente, i dimostrativi rilasciati hanno permesso di affinare il motore di visualizzazione di documenti

IEEE 1599, la ricerca si sta ora concentrando su algoritmi e applicativi per abbattere i tempi di sviluppo dei materiali, che rappresentano uno dei principali motivi che limita il successo del formato.

## 5. IL FUTURO DI IEEE 1599

A distanza di 10 anni dalla promulgazione dello standard, avvenuta nel 2008, l'IEEE ha chiesto di ricostruire il gruppo di lavoro IEEE 1599 al fine di apportare aggiornamenti e migliorie al formato. Questa possibilità di revisione è importante da vari punti di vista: in primo luogo perché permette di rivederne le caratteristiche alla luce delle nuove sfide tecnologiche, in secondo luogo in quanto consente di risolvere alcuni problemi emersi con l'utilizzo di IEEE 1599 in contesti pratici, e infine perché fornisce l'occasione per confrontarsi — nuovamente — con la comunità scientifica nell'ambito del *sound and music computing*.

Tra le principali direzioni da intraprendere, vi è il supporto a formati di notazione musicale altri rispetto alla codifica proprietaria attualmente in uso. Oltre a risolvere dal punto di vista squisitamente teorico l'asimmetria tra i livelli, questo dovrebbe consentire di recuperare un vasto corpus documentale di partiture già disponibili, tra cui gli archivi storici in DARMS, Humdrum e Kern, nonché la produzione in continua evoluzione nei formati MusicXML e MEI, con un occhio di riguardo all'iniziativa di standardizzazione in atto in ambito W3C.<sup>4</sup> Come effetto collaterale, i software di notazione musicale che supportano l'esportazione nei formati citati diventeranno automaticamente strumenti per la creazione del livello logico di IEEE 1599, senza necessità di rilasciare plugin.

L'obiettivo forse più ambizioso della revisione del formato è la generalizzazione della sua struttura multistrato. Sebbene in letteratura sia ampiamente accettata la ripartizione in 6 livelli per l'informazione musicale [13, 14, 15], non si può escludere che in futuro emergano esigenze differenti, ad esempio una maggiore specializzazione dei livelli o la compenetrazione con altri ambiti disciplinari (il cinema, il web, i videogiochi, ecc.). Consentendo all'autore della codifica di definire i propri livelli e le modalità di aggancio con lo spine, il formato guadagnerebbe in generalità e non richiederebbe futuri aggiornamenti. L'obiettivo non è però semplice da realizzare, in quanto implica la definizione di una grammatica apposita, peraltro sufficientemente semplice da non scoraggiare l'utente. Inoltre, tutti gli strumenti che lavorano con IEEE 1599 dovrebbero essere in grado di adattarsi dinamicamente a situazioni non prevedibili a priori, quali la comparsa di un nuovo livello definito dall'utente, al cui interno l'individuazione degli eventi musicali può seguire una logica diversa dagli usuali meccanismi spazio-temporali di sincronizzazione.

Infine, una questione di grande interesse per garantire il futuro successo del formato è la gestione dei DRM, di cui si dovrà occupare un gruppo di lavoro con competenze eterogenee che spaziano dagli aspetti tecnici/tecnologici a quelli prettamente giuridici. Inoltre, dovranno essere adeguatamente rappresentati i soggetti detentori dei contenuti

<sup>3</sup> <http://emipiu.di.unimi.it/>

<sup>4</sup> <https://www.w3.org/community/music-notation/>

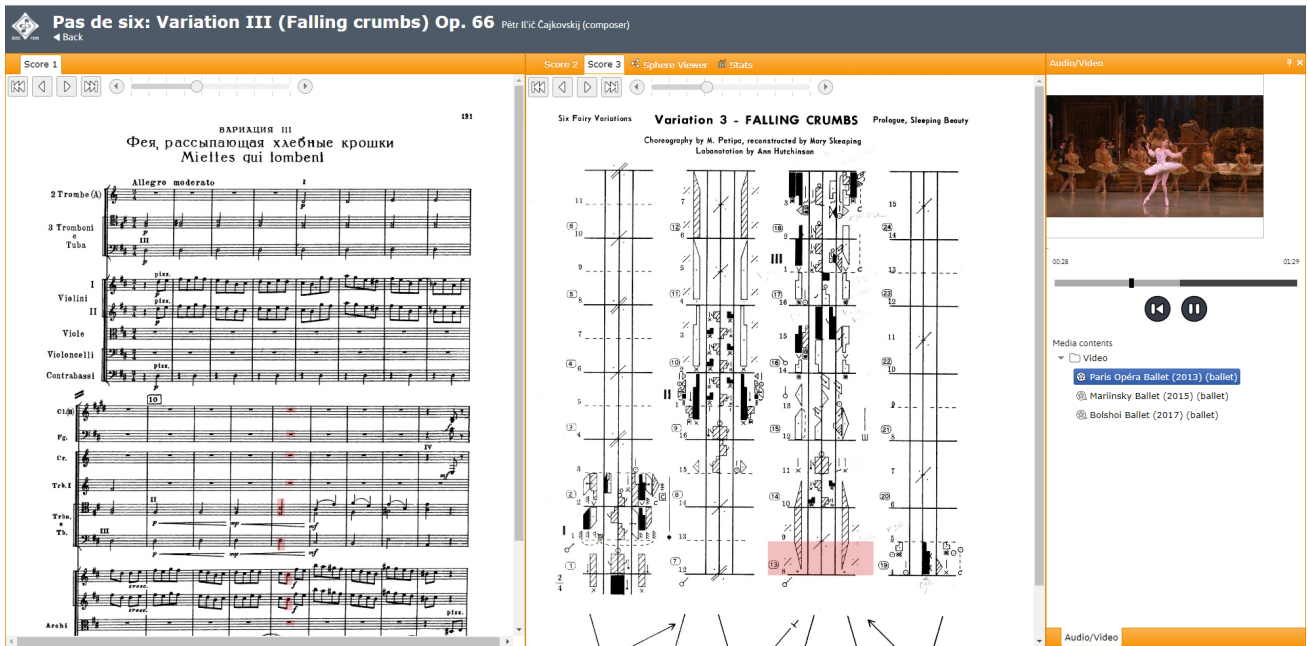


Figura 3. L'interfaccia del nuovo web player per il formato IEEE 1599.

multimediali, al fine di diradare le loro perplessità e coinvolgerli attivamente nel progetto IEEE 1599.

In risposta a queste e altre esigenze, si sta ricostruendo il *working group* IEEE 1599, che – in continuità con la precedente standardizzazione – vede Goffredo Haus nelle vesti di *chair*. Recentemente è stata diffusa dall'IEEE una *call for participation* aperta a tutti gli esperti del settore che vogliono contribuire attivamente alla revisione del formato. La comunità di riferimento da coinvolgere non è solo quella del *sound and music computing*, ma include competenze in ambito tecnologico più ampio, in campo artistico e creativo, nel dominio economico e giuridico, senza peraltro dimenticare il mondo *consumer* e dell'industria.

All'interno di questa cornice si terrà a fine gennaio 2019 un workshop presso la nuova sede di via Celoria 18 del Dipartimento di Informatica dell'Università degli Studi di Milano. Tale iniziativa, chiamata *1st International Workshop on Multilayer Music Representation and Processing (MMRP19)*,<sup>5</sup> fungerà sia da occasione di confronto scientifico sulle tematiche connesse a IEEE 1599, sia da *kickoff meeting* per l'avvio delle attività del ricostituito *working group*.

## 6. BIBLIOGRAFIA

- [1] L. A. Ludovico, "Encoding music information," in *Music Navigation with Symbols and Layers: Toward Content Browsing with IEEE 1599 XML Encoding* (D. Baggi and G. Haus, eds.), pp. 21–36, Hoboken: John Wiley and Sons, 2013.
- [2] A. D'Aguanno, G. M. Haus, and D. A. Mauro, "Feature extraction and synchronization among layers," in *Music Navigation with Symbols and Layers: Toward Content Browsing with IEEE 1599 XML Encoding* (D. Baggi and G. Haus, eds.), pp. 77–95, Hoboken: John Wiley and Sons, 2013.
- [3] G. Haus and M. Longari, *Proceedings of the 1st International Conference on Musical Application using XML (MAX2002)*. IEEE, 2002.
- [4] D. L. Baggi and G. M. Haus, *Music navigation with symbols and layers: Toward content browsing with IEEE 1599 XML encoding*. Hoboken: John Wiley & Sons, 2013.
- [5] A. Baratè, G. Haus, and L. A. Ludovico, "A critical review of the IEEE 1599 standard," *Computer Standards & Interfaces*, vol. 46, pp. 46–51, 2016.
- [6] M. Good *et al.*, "MusicXML: An internet-friendly format for sheet music," in *XML Conference and Expo*, pp. 3–4, 2001.
- [7] P. Roland, "The music encoding initiative (MEI)," in *Proceedings of the First International Conference on Musical Applications Using XML (MAX2002)*, pp. 55–59, IEEE, 2002.
- [8] A. Baratè, G. Haus, L. A. Ludovico, and P. Perlasca, "Managing intellectual property in a music fruition environment - the IEEE 1599 case study," *IEEE MultiMedia*, vol. 23, no. 2, pp. 84–94, 2016.
- [9] A. Baratè, G. Haus, and L. A. Ludovico, "Multilayer formats and the semantic web: A music case study," in *Proceedings of the Audio Mostly 2015 on Interaction With Sound* (C. Dimoulas and G. Kalliris, eds.), pp. 2:1–2:8, ACM New York, NY, USA, 2015.
- [10] A. Baratè and L. A. Ludovico, "IEEE 1599 applications for entertainment and education," in *Music Navigation with Symbols and Layers: Toward Content Browsing with IEEE 1599 XML Encoding* (D. Baggi and

<sup>5</sup> <http://mmrp19.di.unimi.it/>

G. Haus, eds.), pp. 115–132, Hoboken: John Wiley and Sons, 2013.

- [11] A. Baratè, G. Haus, L. A. Ludovico, and D. A. Mauro, “IEEE 1599 for live musical and theatrical performances,” *Journal of Multimedia*, vol. 7, no. 2, pp. 170–178, 2012.
- [12] S. Baldan, L. A. Ludovico, and D. A. Mauro, “Managing multiple media streams in HTML5: the IEEE 1599-2008 case study,” in *Proceedings of the International Conference on Signal Processing and Multimedia Applications (SIGMAP 2011), Seville, Spain* (A. Linares-Barranco and G. A. Tsihrantzis, eds.), pp. 193–199, SCITEPRESS - Science and Technology Publications, Lda., 2011.
- [13] A. Lindsay and W. Kriechbaum, “There’s more than one way to hear it: multiple representations of music in MPEG-7,” *Journal of New Music Research*, vol. 28, no. 4, pp. 364–372, 1999.
- [14] J. Steyn, “Framework for a music markup language,” in *Proceeding of the First International Conference on Musical Application using XML (MAX2002)*, vol. 1060, pp. 22–29, IEEE, 2002.
- [15] G. Haus and M. Longari, “A multi-layered, time-based music description approach based on XML,” *Computer Music Journal*, vol. 29, no. 1, pp. 70–85, 2005.

# FDSDB\_XXth\_CT: una libreria di composizione algoritmica per Open Music e PWGL dedicata alle tecniche compositive del Novecento

Fabio De Sanctis De Benedictis  
 ISSM "P. Mascagni" – Livorno  
 fabio.desanctis@conslit.it

## ABSTRACT

In questo articolo si descrive la libreria di composizione algoritmica FDSDB\_XXth\_CT, per PWGL e per Open Music, dedicata ad alcune delle tecniche compositive del Novecento. È liberamente scaricabile dalla pagina seguente: <https://sites.google.com/site/fdsdbmascagni/code>.<sup>1</sup>

Dopo una rassegna generale, necessariamente non esaustiva, sul tipo di librerie disponibili per PWGL e Open Music, si passa a descrivere il percorso di realizzazione che ha condotto alla creazione della libreria, consequenziale alla pratica analitica e compositiva, e risultato di un lavoro protrattosi negli anni e ancora in corso. Successivamente si illustra in sintesi la struttura della libreria e si elencano le tecniche compositive contemplate, afferenti alla Dodecafonìa, al periodo Seriale e Post-seriale, alla *Pitch-Class Set Theory*, e ad altri procedimenti di elaborazione di altezze e durate. Infine si mostrano alcune *patch* di esempio, seguite da una breve conclusione sugli scopi e l'utilizzo della libreria.

## 1. INTRODUZIONE

Come noto, una libreria *software* è un componente che aggiunge nuove funzionalità ad un programma esistente. FDSDB\_XXth\_CT,<sup>2</sup> come vedremo, contiene numerosi oggetti che svolgono compiti legati a tecniche compositive e compositori del Novecento, che diversamente necessiterebbero di una intera *patch* per essere assolti. Non è questa la sede opportuna per definire in dettaglio i procedimenti necessari alla creazione di una libreria in Open Music o in PWGL, tuttavia è possibile dare qualche indicazione. PWGL contiene alcune istruzioni sulla realizzazione di librerie all'interno della sua documentazione di aiuto. Altrimenti può essere di sussidio la lettura di [1], allegato al *software* o consultabile in rete a questo indirizzo: <https://quod.lib.umich.edu/i/icmc/bbp2372.2003?rgn=full+text>. Per Open Music possiamo rimandare alle pagine

<sup>1</sup> Tutti gli indirizzi Internet riportati in questo articolo sono stati verificati in data 20 settembre 2018.

<sup>2</sup> L'acronimo del nome della libreria contiene, oltre all'evidente riferimento alle iniziali del nome e cognome dell'autore, l'indicazione del ventesimo secolo e le lettere "CT" che stanno per *Composition Techniques*.

Copyright: ©2018 Fabio De Sanctis De Benedictis et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution License 3.0 Unported](https://creativecommons.org/licenses/by/3.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

web della documentazione per sviluppatori, a partire dalla seguente: <http://repmus.ircam.fr/openmusic/sources>. Entrambi i programmi poi contengono delle semplici librerie di esempio; analizzando il codice in esse contenuto e confrontandolo con gli oggetti corrispondenti è possibile adattarlo ai propri scopi. Per un primo apprendimento del codice Lisp ci si è basati su numerose risorse in rete e sui testi cartacei seguenti: [2]; [3]; [4]. Per le risorse Internet possiamo sinteticamente rimandare a quanto riportato in nota. Queste sono state utilizzate come fonte di consultazione, o lette integralmente come nel caso della funzione *iterate*.<sup>3</sup>

Il panorama delle librerie esterne disponibili per PWGL e Open Music è molto vario. Per quanto riguarda PWGL una prima idea è possibile averla visitando la pagina dedicata del sito del programma,<sup>4</sup> ma per reperire le numerose altre librerie disponibili occorre attuare accurate ricerche in rete.<sup>5</sup> La situazione è un po' diversa per Open Music, in quanto è l'Ircam stesso ad elencare le librerie esterne nella pagina web dedicata.<sup>6</sup> Le librerie per PWGL possono essere raggruppate sotto diverse categorie, non sempre univoche, che possiamo riassumere nelle seguenti: traduzione in altri programmi, protocolli o formati come OSC, SDIF, XML, bach; manipolazione ritmica e di altezze; programmazione per vincoli; gestione di programmi per la musica elettronica, come Csound e SuperVP; analisi ed estrazione di caratteristiche da liste, file audio, partiture; strumenti compositivi, probabilistici e di utilità varia. Quelle per Open Music possono essere riassunte in categorie simili e ulteriori, e le loro proprietà sono sommariamente descritte alla pagina web citata. Tuttavia, tranne rarissimi esempi rivolti a uno specifico compositore, o alla composizione spettrale, non esistono casi determinati per quan-

<sup>3</sup> Le risorse di rete utilizzate:  
<http://www.lispworks.com/documentation/HyperSpec/Front/index.htm>;  
<http://clqr.boundp.org/download.html>;  
<http://www.cs.cmu.edu/Groups/AI/html/cltl/cltl2.html>;  
<http://www.mcg.uva.nl/index.html?abstracts/dh-97-g.html&target>;  
<http://lisp.plasticki.com/>;  
<http://jtra.cz/stuff/lisp/sclr/index.html>;  
<https://common-lisp.net/project/iterate/doc/index.html>.

<sup>4</sup> <http://www2.siba.fi/PWGL/downloads.html>. Sulla pagina sono presenti collegamenti alle pagine delle librerie o direttamente all'indirizzo di scaricamento.

<sup>5</sup> Sebbene molte librerie siano disperse sui vari siti dei rispettivi sviluppatori, possiamo tuttavia elencare alcuni indirizzi Internet dai quali è possibile scaricarne diverse: <http://www.sandred.com/Downloads.html>; [http://umanitoba.ca/faculties/music/applied\\_studies/studio\\_flat/StudioFlatDownload.htm](http://umanitoba.ca/faculties/music/applied_studies/studio_flat/StudioFlatDownload.htm). Molto utile sarà poi una ricerca sul sito <https://github.com/> che ospita alcuni programmatori che hanno dedicato i loro sforzi alla creazione di librerie per PWGL.

<sup>6</sup> <http://repmus.ircam.fr/openmusic/libraries>. Anche qui troviamo i collegamenti relativi ad ogni libreria.

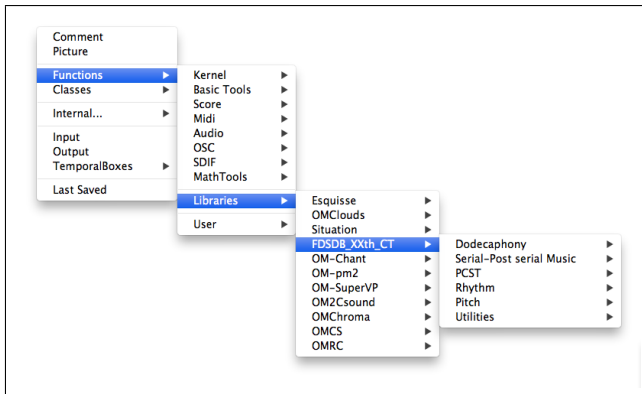


Figure 1. I menu della libreria in Open Music.

to riguarda le tecniche compositive storiche nel loro insieme. Questo uno dei motivi che ha portato alla creazione di FDSDB\_XXth\_CT.

## 2. IL PERCORSO DI REALIZZAZIONE

In un precedente articolo [5] si è accennato alla posizione dell'analisi musicale come disciplina intermedia tra la composizione e l'estetica; questa caratteristica ha reso possibile l'uso di PWGL per realizzare *patch* relative a tecniche compositive di opere del Novecento. Un processo di ingegneria inversa, insomma, che dall'analisi ha portato alla codifica algoritmica delle tecniche compositive dedotte. Questo lavoro in itinere, ancora attivo, ha condotto alla realizzazione di una serie di *patch* in PWGL, alcune delle quali descritte in [5], successivamente tradotte in Open Music. Queste *patch* sono state poi codificate in linguaggio Lisp, verificandone preliminarmente ove opportuno il funzionamento con gli oggetti `lispfunction` di Open Music e `code-box` di PWGL.<sup>7</sup> È stato dunque possibile condensare, per così dire, ogni *patch* in oggetti singoli e determinati. Seguendo poi le istruzioni relative alla realizzazione di librerie in PWGL e in Open Music, adattando dove necessario il codice Lisp, si è arrivati al compimento della libreria qui discussa, prima nella versione per PWGL e poi in quella per Open Music.<sup>8</sup>

## 3. LA STRUTTURA DELLA LIBRERIA

La libreria mostra diversi menu di primo livello (Figura 1): Dodecaphony, Serial-Post serial Music, PCST,<sup>9</sup> Rhythm, Pitch, Utilities.

Ogni menu si articola, dove possibile, nei nomi dei compositori presi in considerazione, e per ogni compositore abbiamo oggetti relativi a sue tecniche compositive.<sup>10</sup> Di-

<sup>7</sup> Gli oggetti citati permettono di inserire in una *patch* un elemento configurato dal codice Lisp scritto al suo interno, e collegabile ad altri componenti della *patch*.

<sup>8</sup> Gli adattamenti di codice e le differenze principali da PWGL a Open Music si sono avuti soprattutto per la gestione di classi relative alla visualizzazione della notazione musicale. Inoltre in PWGL si era ampiamente utilizzata la funzione ricorsiva `iterate` che in Open Music si è dovuta sostituire con la funzione `loop`.

<sup>9</sup> *Pitch-Class Set Theory*.

<sup>10</sup> Si noterà l'assenza di ogni riferimento allo spettralismo. In ambedue i programmi esistono già librerie dedicate a tecniche compositive spettrali. Si è cercato per quanto possibile di evitare duplicati, quindi ci si è

versamente da PWGL, dove ciò non è attuabile, in Open Music ogni categoria di oggetti afferente a un determinato compositore ha come icona una sua rappresentazione, fotografia o dipinto, in modo che anche a livello visivo sia persistente la presenza e la memoria del compositore raffigurato. La libreria ha un aiuto *inline* che, oltre a illustrare il funzionamento degli oggetti, contiene riferimenti bibliografici alle fonti consultate per l'estrazione delle tecniche compositive. Di seguito una descrizione sintetica del contenuto.

Il menu Dodecaphony si articola in sottomenu con i nomi di quattro compositori: Schönberg, Berg, Webern e Babbitt. Per ognuno di questi abbiamo oggetti relativi ad alcune delle loro tecniche compositive:

- Schönberg: le forme seriali canoniche; trasposizioni cromatiche e ordinate della serie, come descritte in [6]; matrici dodecafoniche raffigurate in forma di classi di altezze, numeri di nota MIDI o nomi di nota; generazione casuale di serie dodecafoniche.
- Berg: generazione *random* di una serie omni-intervallare.<sup>11</sup>
- Webern: 22 forme di serie simmetriche, divise principalmente in esacordi e tetracordi, su base cromatica e non, relati tra loro attraverso una delle forme seriali; si veda in proposito [7].
- Babbitt: analisi intervallare dei bicordi di una serie; primo, secondo e terzo sistema di trasposizione per ottenere invarianti; questi sistemi sono discussi in [8].

Il menu Serial-Post serial Music contiene i seguenti sottomenu: Boulez, Carter, Ligeti, Maderna, Manzoni, Messiaen, Xenakis. Le tecniche compositive contemporane sono le seguenti:

- Boulez: matrici numeriche seriali, con o senza applicazione di un modulo, per altezze, durate, ampiezze, utilizzate nella composizione di *Structures Ia* come illustrato in [6]; *mapping* generico; ritmi demoltiplicati, come riportato in [9]; moltiplicazione di accordi secondo [10] e [11]; PDA (*Pitch Duration Association*), con riferimento a [12]; espansione melodica, ancora in [10]; sviluppo di un accordo come in *Dérive*, secondo l'analisi in [13].
- Carter: generazione casuale di accordi simmetrici come quelli utilizzati in *Night Fantasies*, illustrati in [14].
- Ligeti: generazione automatica di una *texture* simile a quella di *Lux Aeterna*, ispirandoci a quanto riportato in [15].
- Maderna: generazione automatica e casuale di quadrati magici di lato 3X3, sino a 12X12.<sup>12</sup>

principalmente concentrati su tecniche per le quali ci risulta non esistano corrispondenti in altre librerie.

<sup>11</sup> Riferimento alla serie di Klein che troviamo nella *Suite lirica* di Berg.

<sup>12</sup> Si vedano in proposito [16] e [17]

- Manzoni: espansione di accordi.<sup>13</sup>
- Messiaen: generazione di figure e frammentazioni ritmiche come attuato per la sinfonia *Turangalila*, secondo le indicazioni in [18]; oggetti per la gestione dei modi a trasposizione limitata; generazione di monodie e polifonie ritmiche, determinate dall'utente o casuali, con ritmi non retrogradabili, tecniche osservabili in [19].
- Xenakis: teoria dei setacci applicata a ritmi e altezze, come descritto in [20] e [21]; matrici numeriche secondo la distribuzione di Poisson come in *Achorripsis*.<sup>14</sup>

Il menu PCST contiene oggetti riferibili a differenti tecniche compositive: determinazione e statistica delle invarianti tra due accordi, nello spazio delle altezze relative e assolute; intersezione delle altezze assolute di un accordo con le trasposizioni di un secondo accordo, secondo le descrizioni contenute in [24]; traduzione di insiemi di classi di altezze in durate; imbricazione, come esemplificato da Forte in [25].

Il menu Rhythm contiene oggetti relativi a: realizzazione di canoni mensurali determinati dall'utente o casuali; rotazione di una serie ritmica (*phasing*) come illustrato in [26]; frammentazione di una serie di durate con o senza l'inserimento di pause; "fenestratura" di una successione di durate;<sup>15</sup> trasformazione di una monodia in polifonia, di nuovo osservabile in [10]; uso proporzionale di serie numeriche; uso di numeri primi o di numeri della serie di Fibonacci; generazione casuale di canoni a mosaico.<sup>16</sup>

Il menu Pitch è composto da oggetti che realizzano l'inversione di una serie di altezze, rispetto a un'asse fisso o mobile, e maschere di tendenza secondo il sistema riportato in [29].

Il menu Utilities contiene alcuni strumenti di utilità che parzialmente differiscono da PWGL a Open Music, principalmente per quanto riguarda la gestione e semplificazione di processi legati alla raffigurazione dei risultati in notazione musicale. Gli strumenti comuni riguardano azioni su liste di sottoliste e concernono la permutazione, la ripetizione e l'ordinamento delle sottoliste.

La libreria, come accennato più sopra, è fornita di un aiuto *inline* che oltre a descrivere gli oggetti dà una spiegazione delle tecniche compositive implicate e una bibliografia di riferimento ove presente. In PWGL vi si accede dal menu Help, invece in Open Music dal menu Help del programma è possibile importare il *tutorial* con gli esempi corrispondenti. In ambedue i programmi selezionando un oggetto e premendo il tasto "D" sulla tastiera del computer si apre una finestra con una spiegazione sintetica. Inoltre in Open Music selezionando l'oggetto e premendo il tasto "T" si apre una *patch* relativa ad esso.

<sup>13</sup> Serie di conferenze tenute da Giacomo Manzoni alla Scuola di Musica di Fiesole dal 26 giugno al 1 luglio 1988.

<sup>14</sup> Si vedano in proposito [22] e [23]

<sup>15</sup> Per "fenestratura" si intende la trasformazione in pause, in questo caso determinata casualmente, di alcune figure di una successione ritmica data.

<sup>16</sup> Sui canoni a mosaico rimandiamo a quanto illustrato in [27] e [28].

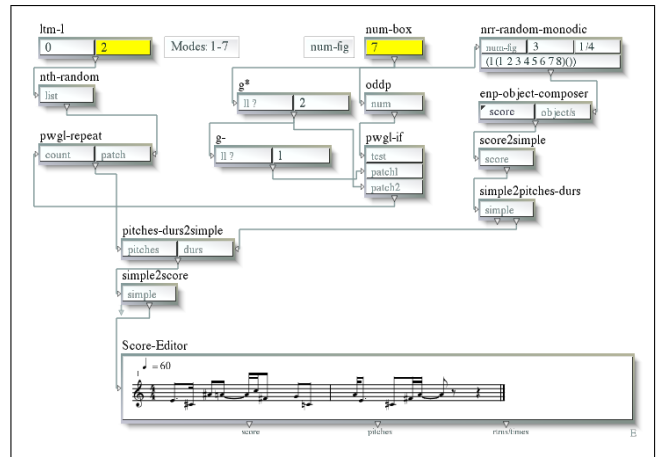


Figure 2. PWGL: generazione casuale di una melodia con ritmo non retrogradabile e altezze di uno dei modi a trasposizione limitata di Messiaen.

#### 4. ALCUNI ESEMPI DIMOSTRATIVI

In questa sezione si illustrano delle *patch*, alcune prese dai *tutorial* della libreria. Ne mostreremo due in PWGL e due in Open Music.

In Figura 2 si osserva la generazione casuale di una melodia le cui altezze sono desunte dal secondo modo a trasposizione limitata di Messiaen, la scala ottatonica, prima trasposizione, e le cui durate si strutturano in un ritmo non retrogradabile generato casualmente. Gli oggetti utilizzati a tale scopo sono rispettivamente *ltn-1* e *nrr-random-monodic*. Ad ogni valutazione<sup>17</sup> dello *Score-Editor* finale si ottiene un nuovo risultato.

Il secondo esempio propone la generazione di una serie di altezze e di una serie di durate alternate con pause, le prime attinenti al profilo dato da una maschera di tendenza. In figura 3 la *patch* corrispondente. Gli oggetti utilizzati sono *mask* per la generazione di altezze secondo la maschera di tendenza, *rtm-frg-with-pauses* che genera una successione ritmica inframezzata da pause, e *chords2score* che trasforma una lista di accordi in una partitura, ogni voce dell'accordo affidata alla stessa parte.

In Figura 4 il risultato della *patch* di figura 3 come visibile nell'oggetto *Score-Editor* risultante.

Passando ad esempi in Open Music, in Figura 5 abbiamo una *patch* che realizza una serie di accordi su durate e altezze desunte da una semplice applicazione della teoria dei setacci in Xenakis. Gli oggetti *sieves-dur* e *sieve-pchs* producono una serie di durate e di altezze; il resto della *patch* si occupa di riunirle in una sequenza di accordi.

La Figura 6 propone una polifonia generata dalla conversione in durate e altezze di un quadrato magico 4X4. Gli oggetti utilizzati sono *latin-square-4X4*, che genera la matrice numerica del quadrato magico, *matrix2voices-rtms*, che converte i numeri in durate, e *matrix2chords*, che effettua la stessa conversio-

<sup>17</sup> Si ricorda che in questi programmi per valutare un risultato si seleziona un oggetto della *patch* e si preme la lettera "V" sulla tastiera alfanumerica dell'elaboratore.

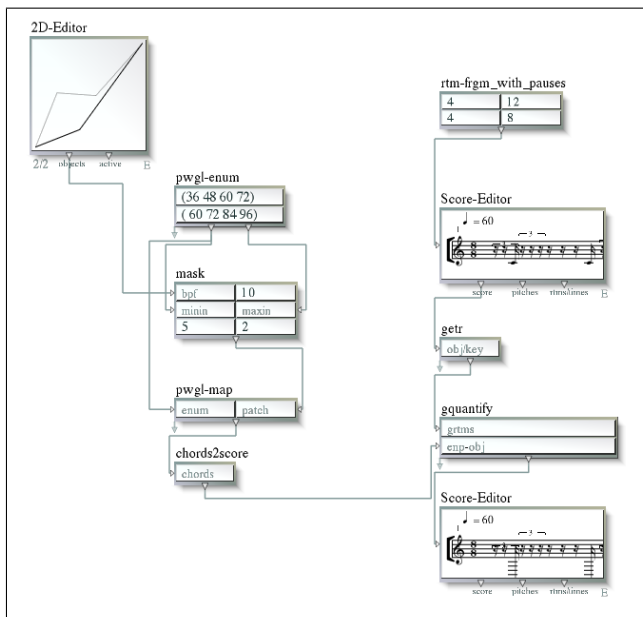


Figure 3. La patch di PWGL che realizza sequenze di altezze che seguono la maschera di tendenza impostata dalle due BPF nell'oggetto 2D-Editor.



Figure 4. Il risultato della patch di Figura 3.

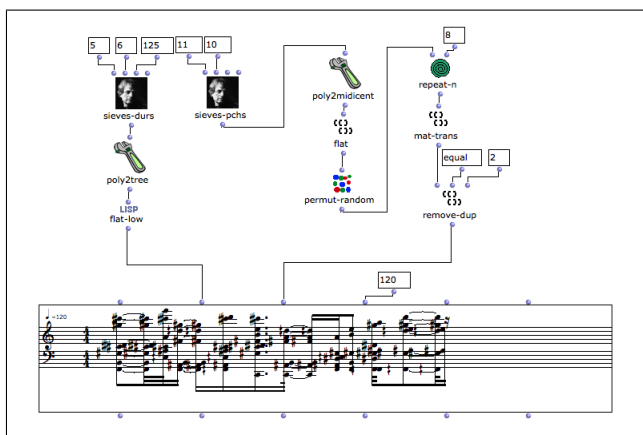


Figure 5. Generazione di una serie di accordi derivati, per altezze e durate, dalla teoria dei setacci di Xenakis.

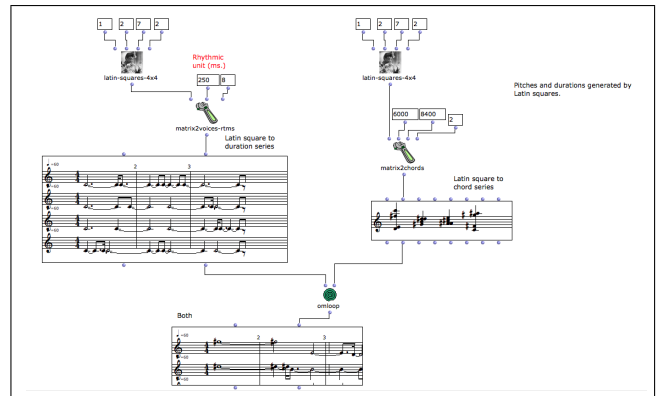


Figure 6. Generazione di una polifonia di altezze e durate da quadrati magici.

ne in altezze. `omloop` riunisce gli elementi generando la polifonia finale.

Gli esempi potrebbero essere molti altri, ma non è questa la sede per trattarne diffusamente: rimandiamo al lettore interessato il compito di esplorare la documentazione e le potenzialità di questa libreria.

### 5. SCOPO E POSSIBILE USO DELLA LIBRERIA

Come specificato, `FDSDB_XXth_CT` è indirizzata alle tecniche compositive del ventesimo secolo. Lo scopo non è solo quello di illustrarle, ma anche quello di diffonderle, di trasmetterle e di preservarle, riconoscendo in esse un valore intrinseco ancora attuale.

Quanto alle applicazioni possibili, a giudizio di chi scrive queste possono andare dalla composizione alla didattica della composizione, nonché trovare spazio nel campo dell'analisi musicale. L'uso ideale, nell'insegnamento, dovrebbe contemplare preliminarmente lo studio delle fonti bibliografiche, l'illustrazione e lo studio delle tecniche compositive, la loro eventuale codifica algoritmica in una patch o in altra forma idonea, e infine l'uso dell'oggetto della libreria che realizza la tecnica presa in considerazione. Tutto ciò affinché l'utilizzo non sia svuotato di significato storico, tecnico ed espressivo, e non diventi un vuoto gioco formale e combinatorio.

Al momento la libreria è utilizzata dallo scrivente, quando opportuno, per lo sviluppo di materiale musicale nelle proprie composizioni. Attualmente si sta lavorando a patch che integrino i processi compositivi con la sintesi per modelli fisici tramite `Modalys` e `Open Music`. `FD-SDB_XXth_CT` non è ancora stata sperimentata appieno nella didattica, tuttavia sono state utilizzate alcune delle patch che hanno poi originato oggetti corrispondenti della libreria. Specificamente si è utilizzato `PWGL` per una Masterclass e due corsi di Tecniche Compositive presso il Conservatorio di Genova, nel corso di Musica Elettronica del M° Doati, e `Open Music` per due corsi di Fondamenti di Composizione II, destinati a studenti di strumento, presso l'ISSM "P. Mascagni", Istituto di titolarità del sottoscritto. In particolare in quest'ultimo caso ci siamo potuti concentrare meno sulla logica dell'algoritmo, e maggiormente sui procedimenti compositivi e formali, permettendo a studen-

ti non compositori di avere un approccio sufficientemente completo e analitico alla musica del Novecento e alla composizione musicale. L'utilizzo degli oggetti della libreria potrà snellire ulteriormente il lavoro, semplificando le *patch*, permettendo così un apprendimento più rapido dei processi e l'apprezzamento immediato del risultato, anche se leggermente a discapito di una profondità di pensiero maggiormente peculiare dei corsi di Composizione.

Da non sottovalutare infine, come accennato, la possibilità di uso con altri oggetti o librerie di PWGL e Open Music, connessi a *software* per la musica elettronica,<sup>18</sup> per la generazione di materiale musicale da utilizzare in questo ambito.

## 6. BIBLIOGRAFIA

- [1] M. Laurson and M. Kuuskankare, "Some Box Design Issues in PWGL," in *Proceedings of International Computer Music Conference*, 2003.
- [2] D. Touretzky, *Common Lisp. A Gentle Introduction to Symbolic Computation*. The Benjamin/Cummings Publishing company, Inc., Redwood city, California, 1990.
- [3] K. Tracton, *Introduzione al Lisp*. Franco Muzzio & C. editore, Padova, 1984.
- [4] G. Gini, M. Gini, and G. Guida, *Lisp. Linguaggio e metodologia di programmazione*. Maggioli editore, Santarcangelo di Romagna (RN), 2007.
- [5] F. De Sanctis De Benedictis, "Dall'analisi musicale alla composizione e formalizzazione algoritmica: esempi applicativi con pwl1," in *Musiche liquide. XX Colloquio di Informatica Musicale*, Roma, 2014.
- [6] G. Ligeti, "Pierre Boulez," in *die Reihe n. 4*, pp. 36–62, 1960.
- [7] M. Babbitt, "Set Structure as Compositional Determinants," in *Journal of Music Theory*, Vol. 5, No. 1, pp. 72–94, 1961.
- [8] M. Babbitt, "Twelve-tone invariants as compositional determinants," in *Musical Quarterly* 46., pp. 246–259, 1955.
- [9] P. Boulez, "Eventualmente," in *Note di apprendistato*, Einaudi, Torino, pp. 135–164, 1968.
- [10] P. Boulez, *Pensare la musica oggi*. Einaudi, Torino, 1979.
- [11] P. Boulez, *Note di apprendistato*. Einaudi, Torino, 1968.
- [12] S. Winick, "Symmetry and Pitch-Duration Associations in Boulez' *Le Marteau sans Maître*," in *Perspectives of New Music*, Vol. 24, No. 2, pp. 280–321, 1986.
- [13] E. Moguillansky, "Continuidad y auto desarrollo en la música de pierre boulez. un análisis de dérive," in P. Cetta (cur.), *Altura-Timbre-Espacio. Cuaderno de Estudio N° 5. IIMCV. Educa. Internet: [http://www.eumus.edu.uy/eme/ensenanza//EP/2008/2008\\_c/documentos/Moguillansky\\_Continuidad.y.auto\\_desarrollo.en.la.musica.de.Boulez.pdf](http://www.eumus.edu.uy/eme/ensenanza//EP/2008/2008_c/documentos/Moguillansky_Continuidad.y.auto_desarrollo.en.la.musica.de.Boulez.pdf)*, pp. 43–52, 2004.
- [14] E. Carter, *Harmony Book*. New York, Carl Fischer, 2002.
- [15] E. Napolitano, "Dalla totalità dispersa del «Requiem» alla coralità senza speranza di «Lux aeterna»," in Restagno, Enzo (ed.), *Ligeti*, Torino EDT, 1989.
- [16] M.J. Böhlen, "Nell'incantesimo del numero. Il Flötenkonzert di Bruno Maderna," in M. Baroni and R. Dalmondo (cur.), *Studi su Bruno Maderna*, Edizioni Suvini Zerboni, Milano, 1985.
- [17] T. Brescia, "Il numero 9 e il magico teorema Luoshu," in *Rivista Punto Zero n. 9*. Internet: <http://www.nexusedizioni.it/it/CT/il-numero-9-e-il-magico-teorema-luoshu-4729>.
- [18] M. Šimundža, "Messiaen's Rhythmical Organizations and Classical Indian Theory of Rhythm (I)," in *International Review of the Aesthetics and Sociology of Music*, Vol. 18, No. 1, pp. 117–144, 1987.
- [19] O. Messiaen, *Technique de mon Langage Musical*. Leduc, Paris, 1944.
- [20] J. Squibb, *An Analytical Approach to the Music of Iannis Xenakis: Studies of Recent Works*. A Dissertation Presented to the Faculty of the Graduate School of Yale University in Candidacy for the degree of Philosophy, 1996.
- [21] C. Soteriou, *Iannis Xenakis: The Analysis of Four Works for Piano Solo*. Thesis submitted to the Charles University in Prague, Faculty of Education, Department of Music Education, 2010/11.
- [22] A. Orcalli, *Fenomenologia della musica sperimentale*. Potenza, Sonus Edizioni Musicali, 1993.
- [23] I. Xenakis, *Musique Formelles*. Paris, La Revue Musicale, 1963.
- [24] R. Morris, *Composition with Pitch-Classes: A Theory of Compositional Design*. Yale University Press, New Haven and London, 1988.
- [25] A. Forte, *The Structure of Atonal Music*. Yale University Press, New Haven and London, 1973.

<sup>18</sup> In PWGL abbiamo PWCsound di Giorgio Zucco per la sintesi in Csound; in Open Music abbiamo le librerie Ircam collegate a Csound (OMChroma, OM2Csound e OMPrisma), quelle che gestiscono SuperVP, pm2, Chant, Spat e Modalys, e librerie esterne tra cui OMCollider, sempre di Zucco, per interagire con SuperCollider, e OM-Sox, che si ricollega all'omonimo programma a linea di comando.

- [26] M. Koenig, *Genesi e forma*. Semar Editore, Roma, 1995.
- [27] T. Johnson, *Tiling the Line in Theory and in Practice*. Editions75. Internet: <http://www.editions75.com/Articles/Tiling%20the%20line%20in%20theory.pdf>.
- [28] T. Johnson, *Tiling the Line in Practice*. Editions75. Internet: [http://repmus.ircam.fr/\\_media/mamux/saisons/saison01-2001-2002/johnson-2002-02-09-b.pdf](http://repmus.ircam.fr/_media/mamux/saisons/saison01-2001-2002/johnson-2002-02-09-b.pdf), 2001-2002.
- [29] A. Cipriani and M. Giri, *Musica Elettronica e Sound Design*. Contemponet, Roma, 2009.

# SampComp: sample-based techniques for algorithmic composition

Andrea Valle

StudiUm/CIRMA - Università di Torino  
andrea.valle@unito.it

## ABSTRACT

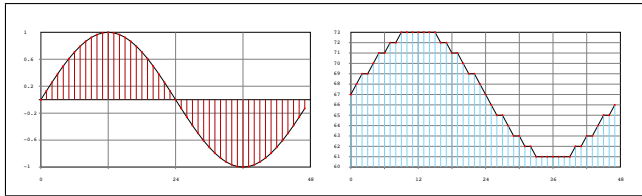
The paper introduces a series of techniques for algorithmic composition in which compositional data are extracted from audio at the sample level. In this way, typical Digital Signal Processing operations can act as tools to create/retrieve music materials. First, in order to describe audio-driven processes in composition, the notion of audioparity is introduced. Secondly, various approaches are discussed that map sample values to composition events. The implementation of these processes is then presented, in particular in relation to music notation generation. Finally, the paper discusses some applications of the same techniques to physical computing.

## 1. INTRODUCTION: AUDIOPARITY, AUDIOPARISM AND TIME SCALING

Algorithmic approaches and techniques to music (regardless of the final medium, whether directly the audio signal, or the musical notation or messages for multimedia objects) are based on a preliminary, general assumption: the presence of some formally defined procedure able to produce a number of representations of musical data. Since the work of Chomsky on formal grammars [1], this kind of approach is typically labelled “generativism”, and in the artistic domain it is so strictly tied to the use of computational resources that historically the terms “computer art” and “generative art” are synonyms [2]. In Chomsky’s approach, the term has a technical meaning (generativism as a formal grammatical theory), but its Latin etymology is indeed linked to human/animal reproduction. In this regard, zoology classically distinguishes between viviparity (the offspring is generated alive), oviparity (the embryonic development takes place in the deposited egg), and ovoviviparity (there are eggs, but they hatch in the mother’s body, and the offspring is born alive). The distinction is interesting because it concerns the ways in which the reproductive behaviour takes place, and is orthogonal to zoological classes (notoriously, the platypus is at the same time mammal and oviparous). In relation to these aspects, the mode of (re-) production and the independence from the class, one could think of a taxonomy of the modes of musical creation, regardless of styles, supports or eras. Consequently, as the suffix “-parity” is used in zoology to indi-

cate reproduction, “audioparity” may be proposed to indicate a composition mode in which musical data originate from sound. A compositional technique is “audioparous” if it defines a projection between a source sound material and an outgoing musical organization. In short, one has audioparity when a composition  $c$  results from the application of an audioparous function  $F_{ap}$  on a sound  $s$ : thus,  $c = F_{ap}(s)$ . Instead, “audioparism” could be used to describe an aesthetic and practical composition attitude centred on audioparity. Such a category does not depend on authors or periods, and, in the same historical context, different composition practices may be more or less audioparous. As an example, French Renaissance composer Clément Janequin has often pursued a pervasive audioparism in many pieces (*La chasse*, *La bataille*, *Les chants des oyseaulx*), explicitly inspired by the transcription of acoustic events. Similarly, in the 20th Century Olivier Messiaen has notoriously integrated the ornithological transcription into many of his works, but often including it into a different composition context [3]. In contemporary music, Spectralism is evidently an area of choice for a pervasive audioparism (see [4]). In his now classic *Les espaces acoustiques* cycle, Gérard Grisey literally re-orchestrates sounds by taking into account data from sonograms: as the French composer said, “we are musicians and our model is sound” (cit. in [5]). On the other side, many composition practices are non-audioparous at all. As an example, Twelve-tone technique is indeed not audioparous, as it organises in various ways an abstract source material, i.e. a 12 pitch-class structure. But the same argument applies to most polyphonic music from 14th to 16th Century. Without discussing further whether the notion of audioparity is relevant or not from a musicological perspective, an interesting point in many audioparous practices, like the ones introduced before, concerns time scale. Messiaen already noted that bird songs must be slowed down in order to match a comfortable time resolution for human ear [3]. In the same pathway, temporal dilatation in relation to the original sound material source is a crucial feature in Grisey’s work. This process of time expansion is not irrelevant, because of non-linearities in time perception. While discussing such an issue in the musical context and following the seminal reflections by Stockhausen [6], Roads notes that, as sound passes from one time scale to another, it crosses perceptual boundaries, as human perception processes each time scale differently [7]. In relation to this, Xenakis has proposed to distinguish four time boundaries (macro/meso/mini/microstructures), that divide time scale into separate perceptual regions, i.e. timbres, notes/events,

Copyright: ©2018 Andrea Valle et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution License 3.0 Unported](https://creativecommons.org/licenses/by/3.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.



**Figure 1.** Sinusoidal signal *src* in a 48-point table (left); requantized version *mld* (right).

rhythms, global evolutions [8]. While these considerations are indeed relevant from a perceptual perspective, still, from the point of view of composition practices, they do not prevent to translate a given feature at a certain time level into another one, as it happens in the aforementioned works by Messiaen or Grisey. The following sections will therefore discuss a set of audiotrans techniques that translate sonic microstructures (specifically, at the sample level) into sonic ministructures (at the event/note level). These techniques can be called SampComp, as they relate the sample level to the composition one.

## 2. SAMPCOMP TECHNIQUES

By providing an abstract unified representation for all data, the numerical domain allows to easily switch between time scales in sound representation. SampComp processes, while in some way related to those digital synthesis techniques that focus on the time-domain representation and operate directly at the sample level (e.g. Xenakis, Koenig, Brün, see [9]), still have with a different goal, that of a time scale expansion. The idea that the signal can act as a control of “notational” parameters is widely used in analog electronic music, for example in Voltage Control technique. As an example, Strange’s 1972 work [10] has investigated the subject by discussing voltage control operations in relation to standard music notation. However, the numerical domain is more pervasive and flexible and allows to precisely map sample data into note ones. As a general case, the following discussion will focus on the generation of a sequence of pitches (i.e. a melody) from a waveform.

The simplest case may take into account a sine wave, as the most basic and usual signal in audio synthesis. The generated digital signal (a cycle) is represented by a vector containing the values in normalised range  $([\pm 1.0])$ , as this is a standard representation, for example widely used for waveform tables in digital oscillators. The resulting digital signal *src*, for an array of 48 elements (an arbitrary size), is plotted in Figure 1, left.

Two basic digital operations can be applied to the *src* signal. The first is resampling. In particular, since the amount of information that describes a signal over time is typically very large if compared to the one describing a melody (e.g. 1 second of audio at CD quality requires 44100 samples), the signal can be resampled at a lower rate than the original. The second is quantization. Similar to sampling, the information of the source signal *src* (in floating point format) has typically a much higher resolution than the one required for pitch description. Quanti-



**Figure 2.** Transcription in music notation of Figure 1 (1 semiquaver per sample).

zation therefore defines the range within which the sample values from *src* are mapped onto pitches. For example, if we consider to map 32 bit audio samples onto an equally-tempered octave (12 pitches), the resolution of the source signal must be reduced to less than 4 bits (as  $2^4 = 16 > 12$ ). These two operations structurally introduce a loss of information and, in fact, a quantity of noise, as they are non-linear. But the exploration of resampling and quantization error is indeed a large part of the interest of the operation. The simplest case consists in resampling by taking a sample every  $n$ . The new sampling rate *srn* will be an integer submultiple of the original sampling *sr*, i.e. the sampling period  $sTn$  will be an integer multiple of  $sT$ . As far as quantization is concerned, a mapping function can be defined that associates the value in the sample domain  $([\pm 1.0])$  to values in a certain range in the integer pitch one by linear interpolation. In Figure 1, on the right, the *src* signal is converted into the new *mld* signal<sup>1</sup>, through a mapping onto the octave that has the middle G as a pivot (hence, in MIDI notation, in the range  $[67 \pm 6]$ , from 61 to 73)<sup>2</sup>, and with  $srn = sr$  (all 48 samples of *src* are taken into account). From this simple example two potentially interesting compositional features emerge:

- to some extent, the sinusoidal waveform of the signal *src* is maintained as the overall form of the melody in the new signal *mld*;
- the quantization to the equally tempered octave introduces a distortion that results in a *terracing* operation, so to speak.

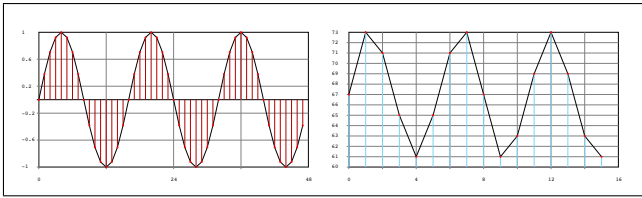
The transcription of *mld* into musical notation (Figure 2) assumes that every sample in *mld* has a duration of one semiquaver.

Expanding further the discussion, the *src* signal in Figure 3, left, again counts 48 samples, but this time 3 cycles of the sine wave are tabulated. The mapping quantization function is the same, but the *mld* signal is obtained by resampling *src* with a sampling period  $sTn = 3 \times sT$  (that is, one sample of *src* is taken every three). As seen in Figure 3, right, the sinusoidal waveform is still visible, but

<sup>1</sup> Sample values are traced in light blue for sake of visibility.

<sup>2</sup> Each float value in the range  $[-1.0, 1.0]$  obtained from *src* is mapped onto the range  $[61, 73]$  by means of  $(\frac{v+1}{2}) \times 12 + 61$ , then rounded and converted to integer. The same mapping will be applied in all following examples for sake of simplicity. Some audio examples are available here:

<http://www.musicaelettronica.it/campcomp-lineamenti-di-musica-audiopara-2/>



**Figure 3.** 3 cycles of a sinusoidal signal (left); requantized version (right).

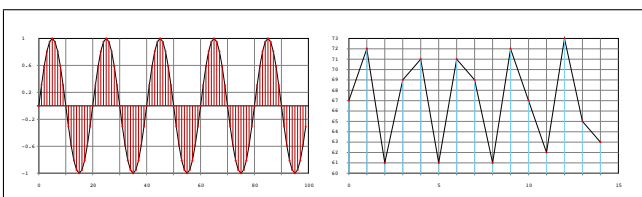


**Figure 4.** Transcription in music notation of Figure 3 (1 semiquaver per sample).

the combined effect of the presence of 3 cycles and sampling/quantization produces a set of asymmetries in the resulting melody. The musical transcription (Figure 4) may decide to assign any duration to any sample of *mld*, but to underline the relationship between *sT* and *sTn* it still assumes the equivalence between a sample of *src* and a semiquaver.

Proceeding further, the following example demonstrates the increase in asymmetry by increasing the number of cycles in *src* (5 in 100 tabulated samples, Figure 5) and re-sampling with  $sTn = 7$  (a sampling that is not yet critical, but certainly important). Although in some way the resulting pitch sequence still bears a sinusoidal shape, the former—as a result of the relationship between resampling and quantization in *mld*—includes only 9 pitches, 4 of which introduced in a sort of final ending tail, while the other 5 are repeated twice (3 times in the case of the C#, MIDI note 61). It can be seen that the resulting pattern is made up of 15 elements. As in the previous examples, the transcription into music notation (Figure 6) maintains the equivalence between source samples and semiquavers: the period  $sTn = 7$  can easily counted as a duration of 7 semiquavers in each note.

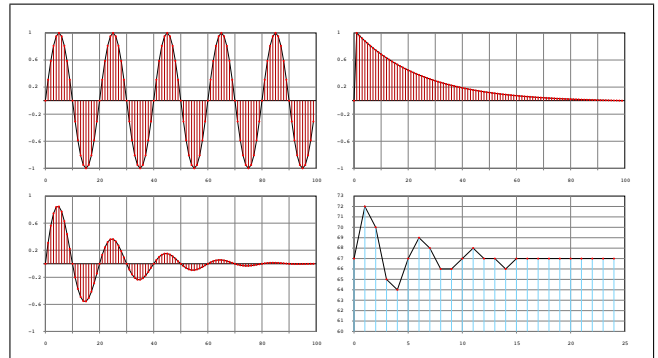
The previous examples have shown that the underlying logic in SampComp is to explore signal time-patterns at the sample level and exploit them at the mini/meso level. This working hypothesis allows then to explore standard Digital Signal Operations (DSP) techniques originally devised for signal generation and manipulation (and well-known in



**Figure 5.** 5 cycles of a sinusoidal signal (left); requantized version (right).



**Figure 6.** Transcription in music notation of Figure 5 (1 semiquaver per sample).



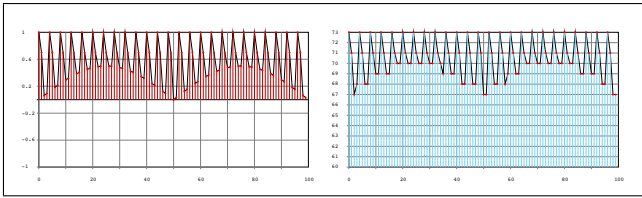
**Figure 7.** Sinusoidal signal (top, left), percussive envelope (top, right), enveloped sinusoid (bottom, left); requantized version (bottom, right).

electronic music) for the purposes of melodic generation. As such operations are defined in the time domain, a basic case is the application of an amplitude envelope. In Figure 7 the *src* signal (bottom, left) is obtained by multiplying 5 cycles of a sinusoid in 100 samples (top, right) by a percussive (i.e. exponentially decreasing) envelope signal (top, left). The melodic contour *mld* (with  $sTn = 4 \times sT$ , 1 value extracted every 4 samples, Figure 7, bottom, right), that result from mapping, clearly shows the relationship between the signal amplitude of *src* and the distribution along the chosen range for pitches. The fading amplitude tail of *src* results in the iteration of the pivot pitch G (MIDI 67). Transcription into musical notation is given in Figure 8.

The application of an envelope can be generalized. The construction of the waveform thus becomes the construction of the melodic contour, in relation to sampling and quantization. In Figure 9, on the left, a 100-point tabulated sinusoid undergoes three operations. The first is a



**Figure 8.** Transcription in music notation of Figure 7 (1 semiquaver per sample).



**Figure 9.** Sinusoidal signal processed by absolute value and impulse replacing (left); requantized version (right).



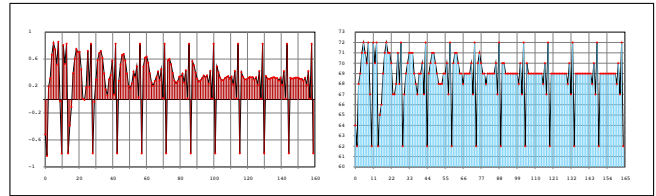
**Figure 10.** Transcription in music notation of Figure 9 (1 semiquaver per sample).

form of digital distortion through the application of the absolute value that reverses the negative half-cycle to positive values (in the audio domain, this process has been proposed by Puckette [11] as a crude but efficient octaver algorithm that introduces distortion). The second is a rescaling of the amplitude (here halved) so that it is compressed in the range  $[0.0, 0.5]$ . Finally, a further operation is applied, iterated with modulo 4 on the samples, which replaces the first and second values of a four-sample block with 1 and 0.7 respectively. Such an operation, while typically numerical, can however be thought of as a form of mixing by replacing, in which samples of the sine-derived wave are regularly replaced by an almost impulsive positive signal. Such a process can be easily generalized, i.e. in terms of mixing of weighted sum of signal vectors to obtain desired shapes (or, with an experimental attitude, to explore emerging melodic shapes from various possible operations). Even in the compressed selected pitch range (an octave) the result in terms of melodic transcription (for  $sTn = sT$ , Figure 10) is interesting as it shows some perceptual features related to grouping. The melody in fact tends to segregate into two streams [12], the pivot top note and the lower “terraced” arpeggio, in a sort of virtual polyphony (the melodic contour is more clearly displayed by Figure 9, right).

In the previous discussion, a fixed resampling rate for  $src$  has been assumed. While fixed sampling frequency is a standard technique in DSP, SampComp techniques may benefit from non standard approaches. In this regard, variable sampling can be applied to a  $src$  signal. In electronics, a Sample & Hold circuit polls the signal, extracts the value, and keeps it until a new polling is asked for [13]. The next example is based on a Sample & Hold algorithm that polls the source signal  $src$  according to a sequence of variable sample intervals instead of a constant rate. Figure 11 is



**Figure 11.** Musical transcription of a variable rate Sample & Hold algorithm applied to one cycle of a sinusoidal signal.



**Figure 12.** A signal obtained by iteratively chaining averaged 15-sample blocks (left); requantized version (right).

the transcription into music notation that results from such a process. Here, a 100-point 1-cycle sinusoid  $src$  signal is sampled along the (cyclic) series  $[3, 5, 7, 2, 2]$ , i.e. the samples 3, 5, 7, 2, 2 are taken from  $src$ , and so on. By counting note durations in semiquavers in Figure 11, the “rhythmic” sampling becomes apparent.

All the previous examples have started from a sinusoid signal as the most basic signal in electronic music. Indeed, the manipulation of signals in the time domain opens up a Pandora’s box of possibilities, far more extensive than the previous preliminary considerations. Further explorations may take into account a large corpus of DSP techniques in order to generate note contour material. In the following, a simple case is thus introduced as an example. An averaging filter is an operation that calculates the value of a sample  $x$  through the average of a certain number  $n$  of nearby samples (before or after  $x$ ) [14]. The recursive application of a filter is at the base of the notorious Karplus-Strong algorithm [15], in which a noisy signal is stored into a table that is rewritten iteratively by applying a median filter (the so-called “recirculating wavetable”). In the following ex-



**Figure 13.** Transcription in music notation of Figure 12 (1 semiquaver per sample).

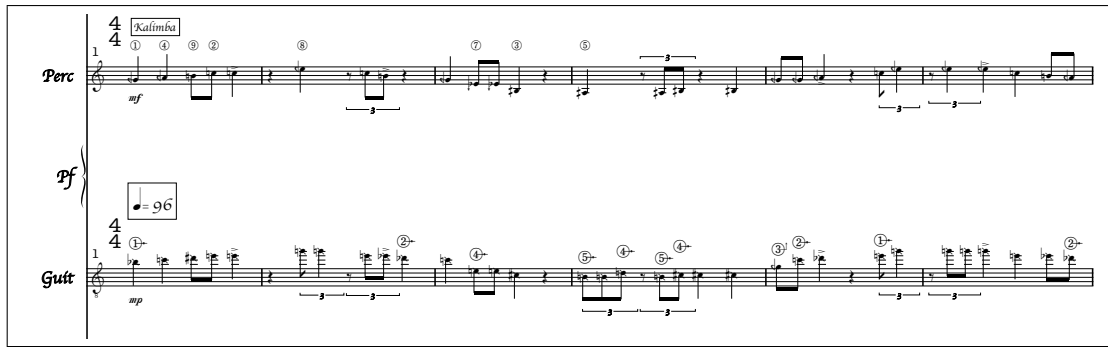


Figure 14. Beginning of *Alatia*, from *Regnum lapideum*.

ample, a *src* signal is purposely constructed in a two-step process. At the initialisation step, a signal  $src_0$  of 15 points is filled with random values. Then, in the second step, the signal  $src_0$  is filtered (in fact, it is a low pass filter) with a step of 3 samples: the value of  $x_n$  is calculated by averaging the samples  $[x_n, \dots, x_{n+2}]$ . The new 15-sample  $src_1$ , that results from filtering, is concatenated with the previous one, then  $src_1$  is filtered again by the same procedure to obtain  $src_2$ , and so on. The filtering and chaining process is repeated 10 times, so that the final *src* signal has a size of 165 samples. As shown in Figure 12, the recursive filtering progressively smoothes the edges of the curve: in the mapped melodic signal *mld*, the initial dispersion of the pitches is thus progressively replaced by a more or less stable curve, centred around a pivot note, as it can be seen in Figure 13. The final result depends crucially on the randomly filled sample block  $src_0$  at the initialization step.

In order to conclude the discussion with a real musical example, Figure 14 shows the beginning of the piece *Alatia*, from the work *Regnum lapideum* by the author and Mauro Lanza, commissioned in 2016 by Ensemble 2e2m (see later). Here the melodic contour is assigned to kalimba and (prepared) guitar. A variable Sample & Hold technique such as the one introduced before has been used to generate the pitch profile. It is easy to recognise a sinusoidal pitch pattern, obtained with a variable (“rhythmic”) polling rate.

### 3. IMPLEMENTATION

The implementation of SampComp techniques by itself does not pose particular problems as the numerical procedures involving the manipulation of small vectors are neither particularly complex nor they require relevant computational resources. All previous examples have been implemented in the SuperCollider environment [16], which features audio synthesis capabilities and an Object-Oriented language to control them. The use of a programming language specifically dedicated to audio is particularly appropriate because the latter comes with a large repertoire of dedicated audio algorithms that can be easily integrated into the workflow. Moreover, through SuperCollider it is possible to sonify the obtained signals and visualise them through graphical interfaces for sake of inspection. A crucial point in algorithmic composition techniques that include musical notation concerns the integration of automatically generated

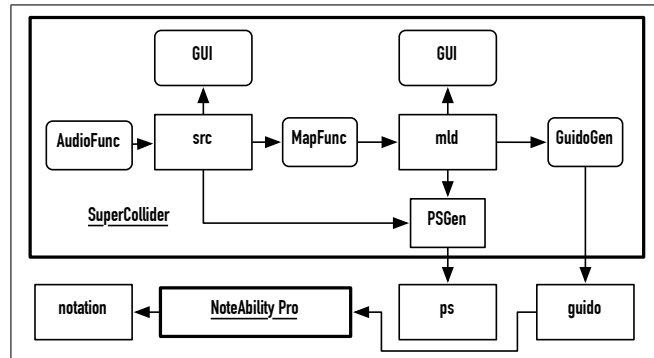
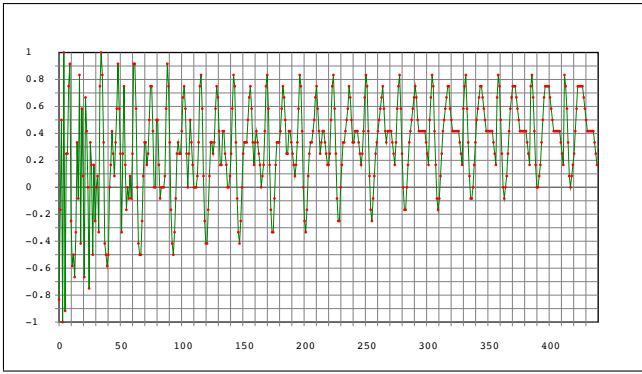


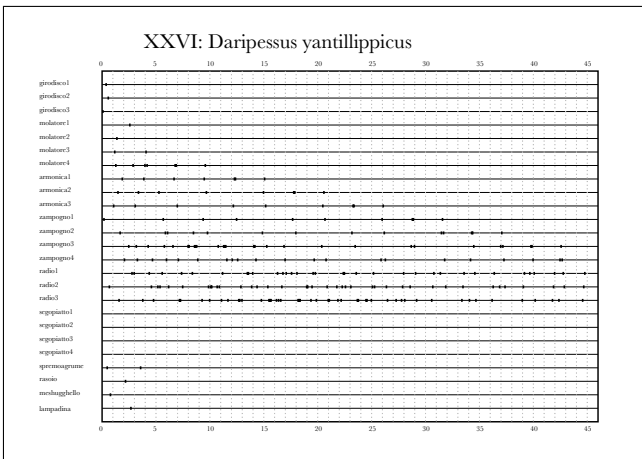
Figure 15. Integrated software pipeline for SampComp and notation generation.

notation in the workflow [17]. In fact, an experimental approach such as SampComp requires that the generation of the notation is integrated into the compositional process in order to rapidly inspect the final results. To this purpose, a light software architecture has been implemented. In Figure 15, the softwares involved are boxed in bold, the software modules are shown in rounded corners, generated data structures and files are enclosed in rectangles. As shown in the Figure 15, SuperCollider implements the generation of the *src* signal (through specialised modules, in Figure labeled as AudioFunc) and the mapping functions (MapFunc) that result in data structures for the representation of pitches (*mld*). SuperCollider also operates as a “glueing” solution. The PSGen module generates text files in the PostScript vector graphic format [18] for the displaying of the generated signals (both *src* and *mld*). The graphic visualisations of signals shown in the previous Figures have been obtained in this way. The integration of music notation into an automated pipeline is not a trivial task. An option could be to generate MIDI files as an intermediate format to be imported into music notation software, the main drawback being the loss of control on music notation typesetting, as the task is left to the importing software. On the other side, MusicXML<sup>3</sup> requires the definition of a large set of typesetting-related parameters. Another option for the generation of musical notation is LilyPond [19], a music engraving software which features a text-based input. Rather, in the present work the musical examples

<sup>3</sup> <https://www.w3.org/2017/12/musicxml131/>



**Figure 16.** Recursively filtered signal used as a source for *Daripessus Yantillipicus*.



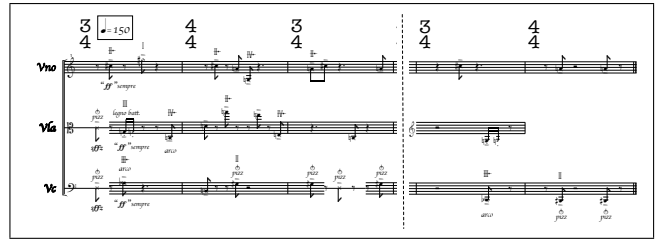
**Figure 17.** Visualisation score for electromechanical devices in *Daripessus Yantillipicus*.

were produced via NoteAbility Pro<sup>4</sup>, an advanced music notation software that includes GUI and playback. NoteAbility Pro can import files written in the compact Guido textual description format [20] and has a much more flexible user frontend compared to the tools available for LilyPond. Thus, a SuperCollider module (GuidoGen in Figure 15) has been implemented, that generates files in the Guido format, so that they can be imported into NoteAbility Pro, played back, eventually re-edited, finally exported as graphic notation files.

#### 4. APPLICATIONS TO OTHER TIME-DOMAIN BEHAVIOURS

Audio DSP techniques provide a vast repertoire of time-patterns that can be used to represent various behaviours on different time scales. In order to extend the previous discussion, this section presents some of the procedures discussed above in relation to the *Systema naturae* cycle for acoustic instruments and electromechanical devices (2013-17), composed by the author and Mauro Lanza [21, 22]. The cycle is made up of four pieces (*Regnum animale*, *Regnum vegetabile*, *Regnum lapideum*, *Fossilia*) and integrates various instrumental ensembles with real-time, computer-

<sup>4</sup> <http://debussy.music.ubc.ca/NoteAbility/>



**Figure 18.** Beginning and ending measures from the instrumental score of *Daripessus Yantillipicus*.

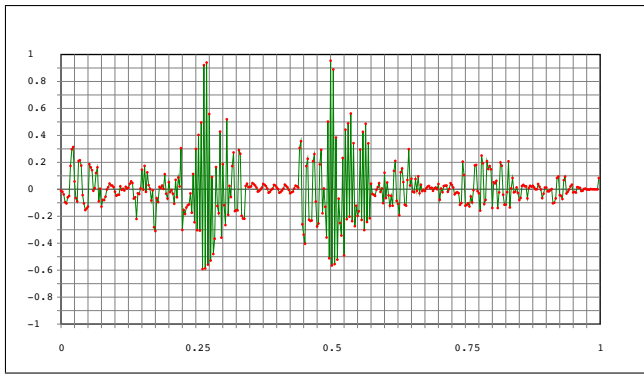
controlled, electromechanical devices, in the perspective of physical computing [23].

In all the four setups of the cycle, the electromechanical devices are obtained from low-cost, scavenged/recycled objects and materials. In order to produce sounds, they are equipped either with small 12 volt DC motors, that provide mechanical energy to scrape, beat, pluck, or blow various objects and materials, or with 12 volt relays, that switch other circuits on/off (e.g. in order to scrape metal plates, they turn on/off modified electric knives operating at 230 volt). While relays work in a binary mode (on/off), DC motors can be fed with a variable amount of current modulated (on a 8 bit scale) by means of Arduino microcontrollers' Pulse Width Modulation (PWM) circuitry [24]. While operating in real-time, in *Systema naturae* these devices follow a pre-composed score that allows a strict synchronisation with human players, as the latter follow a click track. In *Systema naturae*, each of the four works of the cycle is in itself a collection of short pieces. Among these, various have been composed using SampComp techniques, applied both to melodic construction (see the example at the end of Section 2) and to device control.

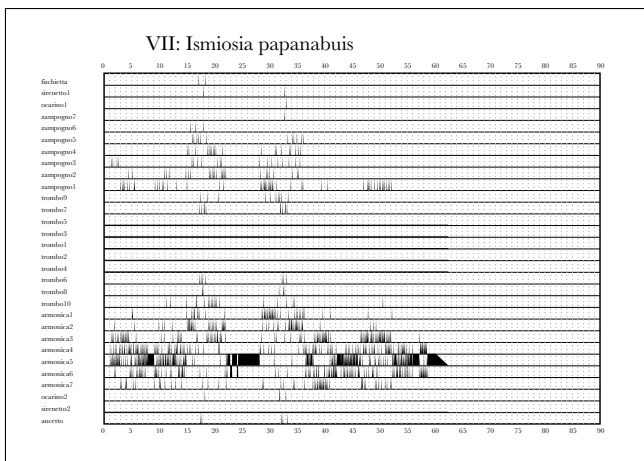
In *Regnum animale*, the piece *Daripessus Yantillipicus* exploits the recursive noise filtering approach discussed above, to define the “melodic” sequence of activation of the devices. This means that values of audio samples are quantized so that they can be associated to the devices, as the latter are ranked in a list. The recursively filtered source signal used to activate devices is shown in Figure 16 (DC offset is not an issue, as *src* will be mapped onto *mld*). The visualisation score is shown in Figure 17<sup>5</sup>. In this case, activation for each device is on/off. The noisy attack in the *src* signal thus results in a dispersion of events, while filtering reduces the range of activated events. In the same piece, such an activation sequence is orchestrated for traditional instruments too. Figure 18 shows the first three and the last two measures of the instrumental parts (separated by a dashed line), so that the filtering from noise-based dispersion of events to a compressed range is apparent.

The same mapping approach (sample values onto ranked devices) is exploited in *Ismiosia Papanabuis* from *Regnum vegetabile*, an example interesting also from a general per-

<sup>5</sup> Score for devices are automatically generated from the composition data structures by exporting PostScript code, following the approach discussed before. From top to bottom, the names of the devices are: girodisco1, girodisco2, girodisco3, molatore1, molatore2, molatore3, molatore4, armonica1, armonica2, armonica3, zampogno1, zampogno2, zampogno3, zampogno4, radio1, radio2, radio3, segopiatto1, segopiatto2, segopiatto3, segopiatto4, spremoaigrume, rasoio, meshuggheho, lampadina.



**Figure 19.** Audio sample (1 second) used as a source signal for *Ismiosia Papanabuis*.

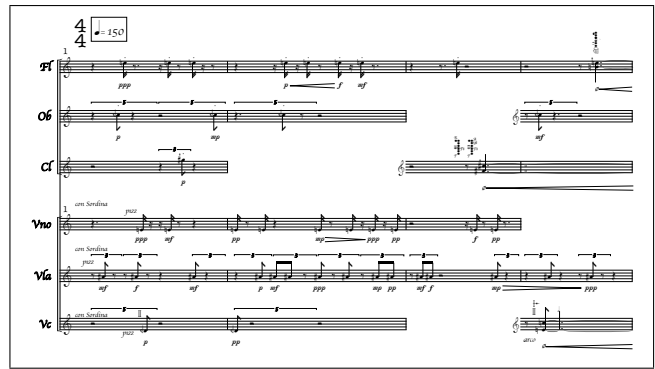


**Figure 20.** Visualisation score for electromechanical devices in *Ismiosia Papanabuis*.

spective. In this case, a short audio fragment is mapped directly onto events, like in *Daripessus*. The 1-second audio sample (Figure 19) is taken from the recording of an electroacoustic improvisation made by the author and Simone Pappalardo, and it has been chosen both as a formal model for composition because of its temporal behaviour and as a homage to Pappalardo, a frequent collaborator of both composers. The only operation performed by MapFunc is the quantization from values in the audio signal to indices representing the selected devices in a ranked list. Figure 20<sup>6</sup> shows the visualisation score for devices. Device activations describe a single melodic contour, following the imported waveform (compare Figure 19 and Figure 20). In the piece, the same principle is applied to acoustic instruments, so that the overall composition is properly a *hochetus* (if attacks, rather than durations, are taken into account), as Figure 21 shows.

As discussed, the electromechanical devices (in case of DC motors) expose as a control interface for their behaviour a single parameter, the 8-bit value addressable for PWM signals. In this sense, they fit the audio signal, as

<sup>6</sup> From top to bottom, devices are: fischieta, sirennetto1, ocarino1, zampogno7, zampogno6, zampogno5, zampogno4, zampogno3, zampogno2, zampogno1, trombo9, trombo7, trombo5, trombo3, trombo1, trombo2, trombo4, trombo6, trombo8, trombo10, armonica1, armonica2, armonica3, armonica4, armonica5, armonica6, armonica7, ocarino2, sirennetto2, anetto.

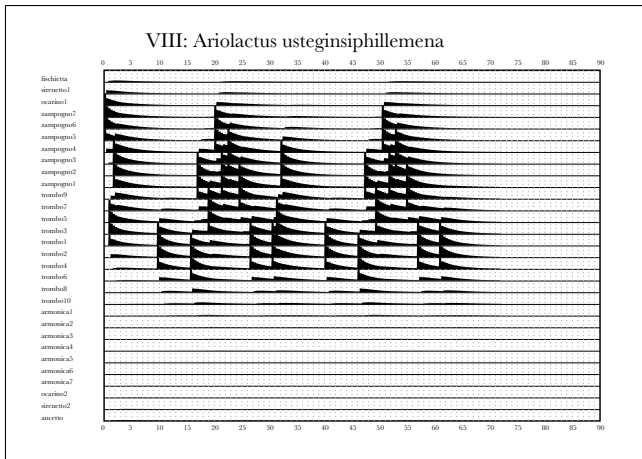


**Figure 21.** Beginning measures from the instrumental score of *Ismiosia Papanabuis*.

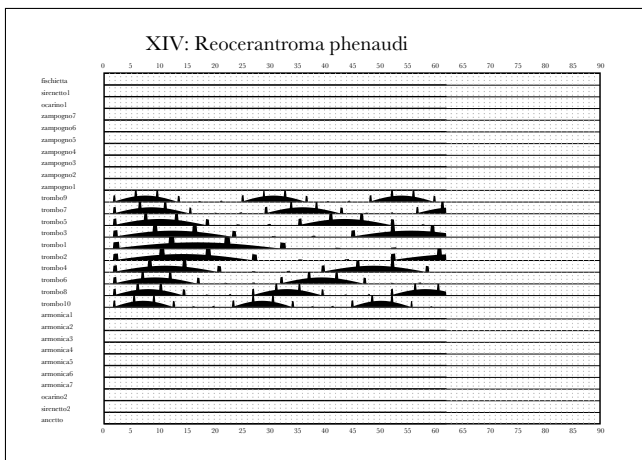
both the PWM and the audio signal are one-dimensional. While in previous examples, all devices were treated as operating in a binary mode, such a PWM control signal can be directly generated via SampComp techniques. The following two examples are from *Regnum vegetabile* too. The average filtering technique is used in the piece *Ariolactus usteginsiphillemena* with a different approach. In the piece, filtering is used to manage the delivery of electric current to group of devices, thus properly implementing a Karplus-Strong-like algorithm that simulates an attack/resonance acoustic model to generate 8-bit PWM signals. Figure 22 reproduces the visualisation score for machines. In this case, the averaging filter is used to directly control the values sent by Arduino's PWM circuitry to devices. Finally, the same approach is at play in the piece *Reocerantroma phenaudi*. PWM signals for devices are obtained by applying the previously introduced technique of the inverted sinusoid with impulse replacement, yielding to activation patterns made up of a continuous behaviour with short spikes (as can be clearly seen in Figure 23).

## 5. CONCLUSIONS

Audioparity can be seen as a musical aesthetics meta-category that groups together many historically and technically different approaches to composition, that share a common interest in sound as a starting point for music creation. Thus, SampComp can be seen as an audioparous approach, provided that sound as a starting point is not considered neither in perceptual nor in acoustic terms, rather in relation to its technological implementation as a digital audio signal. This allows to profit from two mutually related aspects. On one side, the digital encoding of information allows a continuous transition from the sample level to the event one, due to the common numerical format. On the other one, by taking into account digital audio signals, a large corpus of knowledge is available to experimentation from the DSP theoretical domain. In short, the audio signal is taken into account as an abstract cognitive form to describe time patterns. Hence, the possible interest of SampComp techniques, which use can be further expanded and not necessarily limited to pitch contour generation: as an example, to manage the harmonic level or, as discussed in the last section, to generate time forms for generic sound



**Figure 22.** Visualisation score for electromechanical devices in *Ariolactus usteginsiphillemena*.



**Figure 23.** Visualisation score for electromechanical devices in *Reocerantroma phenaudi*.

producing behaviours.

## 6. REFERENCES

[1] N. Chomsky, *Syntactic Structures*. The Hague: Mouton and Co., 1957.

[2] M. A. Boden and E. A. Edmonds, “What is generative art?,” *Digital Creativity*, vol. 20, no. 1-2, pp. 21–46, 2009.

[3] O. Messiaen, *The Techniques of my Musical Language*. Paris: Leduc, 1956.

[4] A. Orcalli, *Fenomenologia della musica sperimentale*. Potenza: Sonus, 1993.

[5] F. Paris, *Le temps de l’écoute. Gerard Grisey ou la beauté des ombres sonores*, ch. L’empreinte du cygne, pp. 51–65. Paris: L’Harmattan, 2004.

[6] K. Stockhausen, “...how time passes,” *die Reihe*, vol. 3, pp. 10–43, 1957.

[7] C. Roads, *Microsound*. Cambridge, Mass. and London: The MIT Press, 2001.

[8] I. Xenakis and R. Brown, “Concerning time,” *Perspectives of New Music*, vol. 27, no. 1, pp. 84–92, 1989.

[9] C. Roads, *The Computer Music Tutorial*. Cambridge, MA, USA: MIT Press, 1996.

[10] A. Strange, *Electronic Music: Systems, Techniques, and Controls*. Dubuque, Iowa: William C Brown Company, 2nd ed., 1983.

[11] M. Puckette, *The Theory and Technique of Electronic Music*. River Edge, NJ, USA: World Scientific Publishing Co., Inc., 2007.

[12] A. Bregman, *Auditory Scene Analysis. The Perceptual Organization of Sound*. Cambridge, Mass. and London: The MIT Press, 1990.

[13] P. Horowitz and W. Hill, *The Art of Electronics*. New York, NY, USA: Cambridge University Press, 3rd (2015) ed., 1989.

[14] D. Rocchesso, *Introduction to Sound Processing*. Firenze: Mondo Estremo, 2003.

[15] K. Karplus and A. Strong, “Digital synthesis of plucked-string and drum timbres,” *Computer Music Journal*, vol. 7, no. 2, pp. 43–55, 1983.

[16] S. Wilson, D. Cottle, and N. Collins, eds., *The SuperCollider Book*. Cambridge, Mass.: The MIT Press, 2011.

[17] A. Valle, “Integrated Algorithmic Composition. Fluid Systems for including notation in music composition cycle,” in *NIME 2008: Proceedings*, pp. 253–256, 2008.

[18] Adobe, *PostScript Language Reference*. Reading, Mass.: Addison-Wesley, 3rd ed., 1999.

[19] H.-W. Nienhuys and J. Nieuwenhuizen, “LilyPond, a system for music engraving,” in *Proceeding of the XIV CIM 2003*, (Firenze), pp. 167–172, 2003.

[20] H. H. Hoos, K. A. Hamel, K. Renz, and J. Kilian, “The GUIDO Notation Format – A Novel Approach for Adequately Representing Score-Level Music,” in *ICMC’98 Proceedings*, pp. 451–454, 1998.

[21] P. Roullier, ed., *Mauro Lanza et Andrea Valle: Systema naturae*. Champigny-sur-Marne: Ensemble 2e2m, 2016.

[22] A. Valle and M. Lanza, “Systema naturae: shared practices between physical computing and algorithmic composition,” in *Proceedings of the 14th Sound and Music Computing Conference* (J. P. Tapio Lokki and V. Välimäki, eds.), (Espoo), pp. 391–398, Aalto University, Aalto University, 2017.

[23] D. O’Sullivan and T. Igoe, *Physical Computing. Sensing and Controlling the Physical World with Computers*. Boston, Mass.: Course Technology, 2004.

[24] M. Banzi, *Getting started with Arduino*. O’Reilly, 2009.

# IL PROTOTIPO “SCUBA”: UN INDUMENTO AUMENTATO PER PROGETTARE L’INTERAZIONE SONORA CON I VESTITI

Giorgio Klauer

Conservatorio “B. Marcello”, Venezia

giorgio@klauer.it

## ABSTRACT

Viene presentato un prototipo di indumento concepito quale strumento per progettare il suono e l’interazione sonora con i vestiti. Le caratteristiche salienti del dispositivo (realizzazione in neoprene, sensori e altoparlanti applicati, microcomputer incorporato, sistema di gestione remota) conferiscono qualità complesse in fatto di *interaction design* ed *embodiment*. Vengono descritte alcune esperienze di utilizzo e, visto l’ambito in cui si svolge il progetto, si effettuano alcune considerazioni in merito al posizionamento di questo tipo di ricerca.

## 1. INTRODUZIONE

Il progetto è la prosecuzione di un’iniziativa dedicata alla progettazione di vestiti aumentati mediante l’interazione sonora in contesti funzionali realizzando prototipi di carta e cartone con tecnologia *embedded*.<sup>1</sup> Svoltasi presso il Conservatorio di Padova (29 febbraio-19 marzo 2016), essa aveva tra gli scopi il trasferimento metodologico da e verso l’ambito della musica elettroacustica utilizzando una modalità laboratoriale partecipativa [1].

Al termine dell’attività si erano evidenziati gli accorgimenti e le opportunità di sviluppo che seguono:

- approfondire il ruolo dell’artista nel processo decisionale del design partecipativo;
- raccogliere sistematicamente dati e verbalizzazioni per consentire una autentica valutazione;
- scegliere preventivamente il software e gli strumenti per l’interazione sonora in modo da equilibrare il contributo dei singoli partecipanti;
- favorire l’azione simultanea dei partecipanti sul sistema di interazione;
- concentrare le risorse, specialmente sonore, per orientare il confronto e la discussione.

Era emersa inoltre l’opportunità di realizzare un prototipo in tessuto, anche sulla base del positivo riscontro

<sup>1</sup> [http://www.sampl-lab.org/it/index.php/Sonic\\_Interaction\\_Design\\_for\\_Paper\\_Artworks\\_and\\_Artifacts](http://www.sampl-lab.org/it/index.php/Sonic_Interaction_Design_for_Paper_Artworks_and_Artifacts)

Copyright: ©2018 Giorgio Klauer. This is an open-access article distributed under the terms of the [Creative Commons Attribution License 3.0 Unported](https://creativecommons.org/licenses/by/3.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

della messa in atto, in via tentativa, di uno scenario in un contesto di utilizzo reale [2].

Le problematiche affrontate nel prosieguo del progetto sono state dunque:

1. la realizzazione di un prototipo di indumento in tessuto, utile a sviluppare con maggiore naturalezza e realismo nell’indossarlo le interazioni sonore connesse a posture, movimenti e gesti;
2. la predisposizione di un ambiente informatico che consentisse l’intervento attivo di più partecipanti simultaneamente, ossia, rendere accessibile in rete il sistema di interazione in modo da poterlo modificare durante il suo utilizzo.

## 2. PROTOTIPO “SCUBA”

Il prototipo “scuba” è un indumento interattivo che incorpora sensori, altoparlanti e un microcomputer, finalizzato a progettare e valutare l’interazione sonora con i vestiti in una varietà di contesti di utilizzo dove la comunicazione verbale viene sostituita o viene accompagnata da una comunicazione gestuale e sonora.

Più che un oggetto funzionale singolare o un vero e proprio *smart cloth*, l’indumento viene considerato a sua volta come una piattaforma di prototipazione. Rispetto alla concezione sartoriale e alla sperimentazione in fatto di *electronic textile*, quindi, non si è preso in considerazione il confezionamento di sensori e altoparlanti come elementi tessuti e cuciti, è stata preferita la fattura con tessuto in policloroprene, sono stati utilizzati componenti elettronici *ready-made* e anziché appoggiarsi a uno smartphone e alla tecnologia Bluetooth/NFC/RFID (tra le *best practice* correnti) è stato utilizzato un single board computer.

L’aspetto esteriore è concepito in modo neutro; pochi elementi decorativi non funzionali permettono eventualmente, in uno scenario dimostrativo, di distogliere l’attenzione dagli aspetti più tecnici della ricerca. Nello specifico, la fattura è derivata da due fra i prototipi output del workshop citato in precedenza, denominati rispettivamente *fashion show* e *theater usher*, di cui si riprendono alcune *feature* astraendone però le funzionalità dal contesto di utilizzo previsto.

Attraverso il software appositamente programmato è possibile correlare a una varietà di atteggiamenti corporei e situazioni, la riproduzione di suoni preregistrati e il controllo parametrico di processi di sintesi ed elaborazione sonora.

## 2.1 L'indumento

Il policloroprene è utilizzato prevalentemente nell'abbigliamento sportivo ma possedendo caratteristiche quali ottima elasticità, morbidezza, volume, robustezza e facilità alla lavorazione, trova impiego in vari settori del *fashion design*.

Nel contesto del progetto, i requisiti materiali del neoprene più utili al confezionamento di un primo paio di esemplari sembravano la resistenza allo sfilacciamento e la robustezza comparata allo spessore, il che consentiva di applicare componentistica *ready-made*.

Il confezionamento di due giubbini è stato effettuato a cura della sarta creativa e maker Ilaria Papis.<sup>2</sup> I due esemplari di colore *bluette* si distinguono nell'applicazione di una fodera interna ovvero di nastri di passamaneria per isolare i fili conduttivi, nell'aggiunta di bande decorative sulle maniche di colore rispettivamente giallo e bruno, e nella taglia *medium* e *large*, oltre che per alcuni altri dettagli.



Figura 1. Valutazioni preliminari al cartamodello.

Il filo conduttivo intrecciato, selezionato interamente metallico per diminuire la resistività del circuito, è stato cucito a macchina sul rovescio del tessuto in modo che esternamente il secondo filo, di colore blu, si integrasse perfettamente nella superficie. Lo spessore del filo conduttivo ha causato non poche difficoltà nell'effettuare anche le più semplici cuciture diritte.<sup>3</sup>

Le cuciture hanno seguito un tracciato a ventaglio originando posteriormente all'altezza della vita. Qui è stato possibile assicurare con due elastici, sul lato interno, il microcomputer. Le sue dimensioni determinavano un inellegante rigonfiamento che potrà essere corretto, in futuro, usando un single board computer più compatto.

I sensori sono stati inseriti in piccole tasche ricavate internamente o esternamente.

Le connessioni ai capi del filo conduttivo, rispettivamente, con i sensori e il microcomputer sono state realizzate con cavetti jumper DuPont in modo da facilitare la sostituzione della componentistica. Questa caratteristica si è rivelata molto vantaggiosa nonostante l'ulteriore ingombro.

La scelta preliminare dei sensori, potenzialmente ricollocabili e sostituibili, ha seguito il criterio della varietà:

<sup>2</sup> <https://www.facebook.com/TailorSid>

<sup>3</sup> Per le problematiche correlate alla cucitura degli *electronic textile* si veda [3].

- gomito: sensore di flessione 112mm<sup>4</sup>
- spalla (regione cervicale): sensore di flessione 77mm<sup>5</sup>
- polso sinistro (parte esterna): sensore di distanza a ultrasuoni HC-SR04
- polso sinistro (parte interna): accelerometro-giroscopio-magnetometro LSM9DS0
- polso destro (parte esterna): sensore di colore TCS34725
- polso destro (parte interna): microfono amplificato MAX4466
- petto (lato sinistro): potenziometro flessibile circolare<sup>6</sup>
- petto (lato destro): potenziometro flessibile diritto<sup>7</sup>
- petto (lato destro): sensore di forza 18mm<sup>8</sup>
- vita (lato sinistro): buzzer PB-12N23P-85Q
- girocollo (parte anteriore): coppia di altoparlanti KP5083SP1-5844 50mm

I resistori per il partizionamento della tensione sono stati saldati ai piedini dei sensori analogici; i valori di resistività utili a massimizzare l'escursione elettrica sono stati individuati sperimentalmente.

Anche il modello degli altoparlanti è stato individuato sperimentalmente, mediando tra le esigenze in fatto di pressione acustica, gamma di frequenze, peso e dimensioni.

Si è optato per l'utilizzo di un microcomputer Beagle-Bone Black in combinazione alla scheda Bela<sup>9</sup> onde riservarsi la possibilità di sostituire la tipologia di alcuni sensori e sfruttarne il segnale in sincrono con l'audio, o di inserire al posto di un paio di sensori ulteriori altoparlanti per una sonificazione multicanale. Al momento non sono stati fatti tentativi al riguardo; la diffusione è monofonica e utilizza l'amplificatore incorporato LM4876.<sup>10</sup>

Come applicativo principale per l'elaborazione del suono è stata usata la versione specifica per la scheda Bela di SuperCollider.<sup>11</sup> I valori dei sensori analogici, del sensore di distanza e del buzzer vi sono rappresentati come *stream* audio campionati a 22kHz. I valori dei sensori TCS34725 e LSM9DS0 connessi ai bus I<sup>2</sup>C vengono letti e trasmessi in messaggi OSC da uno script python per essere trasformati in ulteriori flussi.

L'alimentazione è fornita da una batteria a 5V che potrebbe essere cucita sul vestito ma, per ragioni pratiche, è stata finora inserita nella tasca dei pantaloni. Il consumo è minimo, rendendo un *power bank* da 2600mAh sufficiente alle esigenze iniziali.

<sup>4</sup> <https://www.adafruit.com/product/182>

<sup>5</sup> <https://www.adafruit.com/product/1070>

<sup>6</sup> <https://www.adafruit.com/product/1069>

<sup>7</sup> <https://www.adafruit.com/product/178>

<sup>8</sup> <https://www.adafruit.com/product/166>

<sup>9</sup> <https://github.com/BelaPlatform/Bela/wiki>

<sup>10</sup> Si può prendere pertanto in considerazione l'utilizzo di un PocketBeagle in combinazione alla scheda Bela Mini (<https://blog.bela.io/2018/02/22/bela-mini-launch>) e un amplificatore miniaturizzato, per ridurre l'ingombro, o di un Raspberry Pi (per es. modello 1 A+) con una scheda Arduino compatta (per es. Micro o Nano) oppure una scheda convertitore A/D.

<sup>11</sup> <https://github.com/sensestage/supercollider>

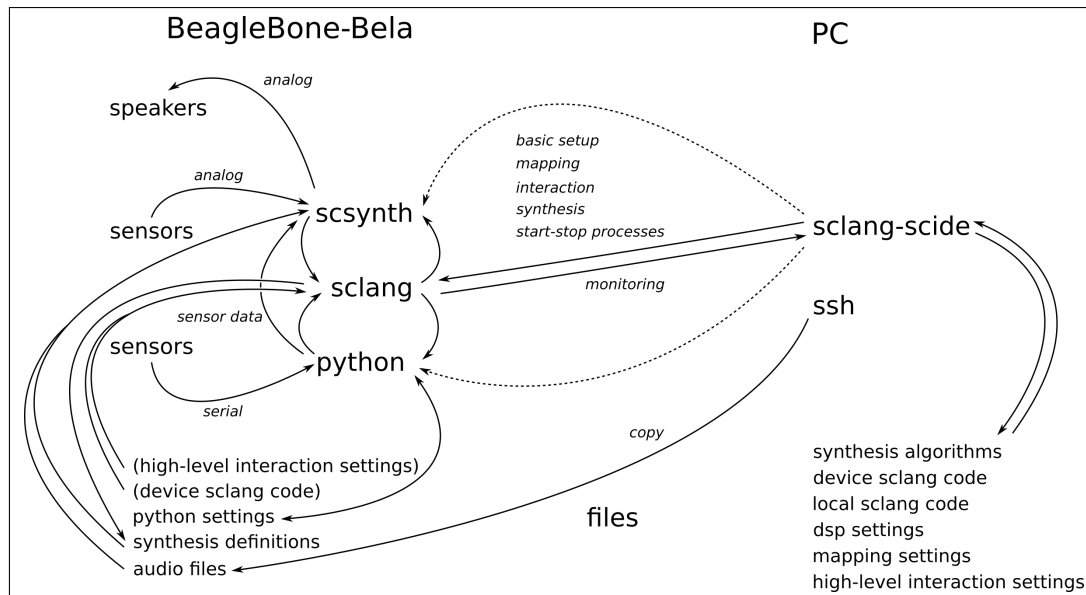


Figura 2. Schema del sistema per il design dell'interazione sonora.

## 2.2 Il sistema

L'attuale sistema è concepito secondo un'architettura client-server tale da permettere, nel design sonoro, l'intervento collaborativo di più utenti da postazioni in rete, nel momento in cui il giubbono è indossato e attivo.

Come soluzione ideale è stata individuata la combinazione del software scsynth e di un webservice python a richiamo degli script necessari a gestire l'interazione. Tuttavia, considerato lo stato non ancora avanzato del progetto e la provenienza degli utenti quasi esclusivamente dall'ambito della musica elettroacustica e delle tecnologie del suono, si è optato per un approccio distribuito combinando l'utilizzo del software slang sulle postazioni e l'insieme di scsynth, slang, python e alcune *utility* di sistema richiamate da script bash a bordo del microcomputer. Questa soluzione ha di fatto agevolato lo sviluppo del dispositivo di interazione sonora ed evidenziato l'implementare un webservice come una risorsa da problematizzare rispetto alla flessibilità imposta da varianti di design che si potrebbero prospettare in futuro.<sup>12</sup>

All'accensione del dispositivo il sistema è "vuoto" in attesa di istruzioni impartite via rete dalle postazioni remote. L'intero processo di inizializzazione, quindi, è avviato da una postazione inviando via OSC al dispositivo tutto il codice slang necessario al funzionamento.<sup>13</sup> Per agevolare il debug nella composizione delle istruzioni per il design dell'interazione, si è implementato un server OSC anche per python e fatto in modo che scsynth, slang e python utilizzassero la medesima sintassi per le comunicazioni tra i software all'interno del microcomputer e dalle postazioni (Figura 2).

La progettazione dell'interazione sonora dalle postazioni si effettua per mezzo di una rudimentale interfaccia gra-

fica realizzata in slang-scide, disposta su due finestre. La prima finestra comprende un set di istruzioni per verificare il corretto funzionamento dei sensori e i valori dei bus audio e di controllo, oltre a popolare i buffer audio, trasferire file etc.

La seconda finestra (Figura 3) permette, nella parte sinistra, di riportare i valori dei singoli sensori in altro ambito e scala dando luogo a vari set di bus di controllo. Nella parte destra vengono richiamati ed editati gli algoritmi di elaborazione sonora. Gli argomenti dell'algoritmo di elaborazione sonora possono essere inseriti singolarmente oppure ricavati da uno o più bus di controllo miscelati o sommati tra loro. Ciascun algoritmo può essere istanziato più volte, nei limiti della potenza di calcolo del microcomputer e dell'implementazione per ora subottimale degli algoritmi generanti i flussi di controllo. L'azione sull'interfaccia grafica determina un aggiornamento delle istanze attive e, in ogni caso, dà luogo alla registrazione in locale di settaggi temporanei. L'utente può salvare e richiamare settaggi di alto livello che includono i vari settaggi di livello più basso relativi alla singola istanza.

Allo stato attuale si è valutato un numero minimo di algoritmi di sintesi e riproduzione/elaborazione di suoni campionati con parametri di basso livello. Di conseguenza, la questione se implementare algoritmi con parametri di più alto livello e denominazioni eloquenti e funzionali è stata rimandata a uno stadio di sviluppo più avanzato del progetto, anche in considerazione del fatto che tale implementazione costituisce una ricerca a se stante per cui esistono modelli di riferimento [4].<sup>14</sup> È senz'altro pensabile un utilizzo degli algoritmi su due livelli, il più alto dei quali sia concepito come un *plug-in* di uso intuitivo/qualitativo dedicato a un'utenza specifica (per es. i designer).

Alcuni video realizzati durante la lavorazione del prototipo per offrire idee più precise sul risultato sonoro sono

<sup>12</sup> Si confronti ad esempio la soluzione di interfaccia browser realizzata dagli sviluppatori di Bela, <https://github.com/BelaPlatform/Bela/wiki/Bela-IDE>.

<sup>13</sup> In aggiunta è prevista la possibilità di eseguire un auto-setup all'avvio in base a file di configurazione di alto livello.

<sup>14</sup> <http://soundobject.org/SDT>

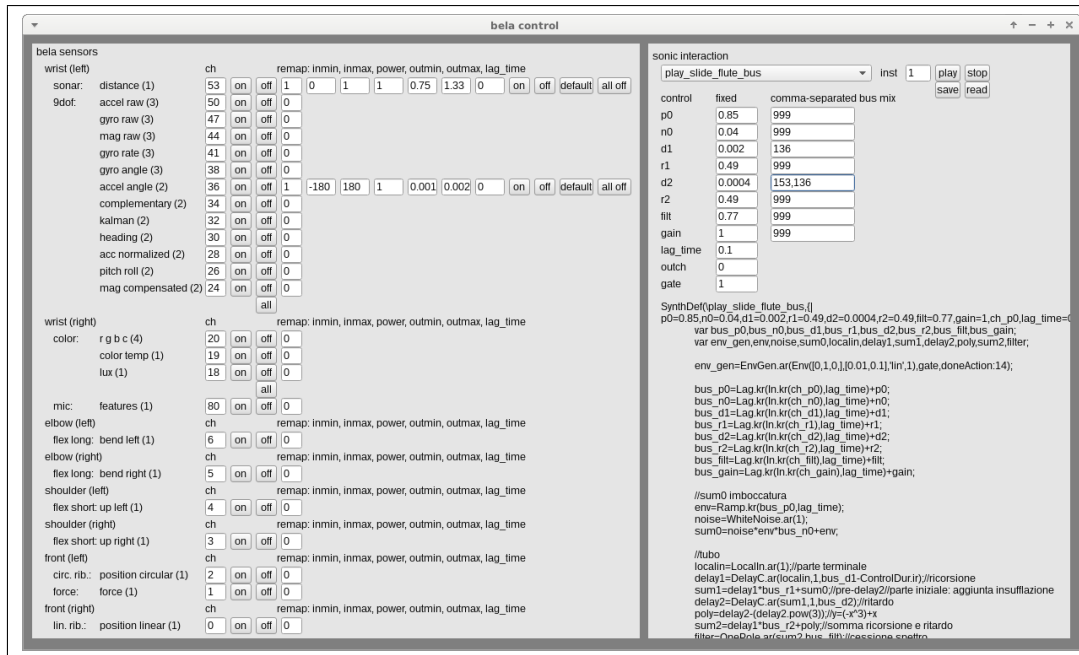


Figura 3. Rudimentale GUI di gestione dell’interazione sonora (parte di progettazione).

disponibili al sito <http://klauer.it/index.php?res=sonification/sound%20cloth/demo>.

### 3. UTILIZZI

#### 3.1 Workshop

Il prototipo “scuba” è stato presentato inizialmente nell’ambito di un secondo workshop svoltosi presso il Conservatorio di Padova (31 maggio-3 giugno 2017).<sup>15</sup>

Il workshop aveva lo scopo di una prima valutazione dell’indumento interattivo. Visto il contesto di svolgimento, si proponeva specialmente di esplorare le categorie della qualità e dell’originalità sonora ricavabili per mezzo del dispositivo.

Componeva il workshop una prima parte volta a illustrarne le caratteristiche tecniche e una seconda parte finalizzata alla creazione di interazioni sonore alternative ed esteticamente valide per possibili scenari d’uso del capo di abbigliamento, inclusa la performance artistica.

##### 3.1.1 Svolgimento

Dopo una premessa in merito a *wearable technology*, *electronic textile*, *smart cloth* e la menzione degli ambiti di utilizzo in cui l’applicazione delle tecnologie indossabili agli indumenti è più avanzata, i partecipanti venivano introdotti al prototipo, soffermandosi sui problemi aperti correlati agli sviluppi delle tecnologie indossabili così come presentati in letteratura.

La seconda parte, che si svolgeva nell’arco di due giornate non consecutive, era volta alla messa in atto di alcune possibilità di interazione sonora.

<sup>15</sup> [http://www.sampl-lab.org/it/index.php/Sound\\_design\\_e\\_interazione\\_sonora\\_per\\_i\\_vestiti\\_aumentati](http://www.sampl-lab.org/it/index.php/Sound_design_e_interazione_sonora_per_i_vestiti_aumentati)

Dapprima ci si è soffermati sulle modalità di attivazione di campioni sonori in base ad ambiti e scale di azione dei sensori e quindi ai movimenti e potenzialmente ai gesti; successivamente, impiegando un algoritmo di sintesi per modello fisico del suono del flauto [5] si valutavano esempi di interazione continua determinando tentativamente l’apporto dei sensori di interesse verso i singoli controlli parametrici.

Per acquisire familiarità con l’indumento interattivo, si effettuavano delle improvvisazioni: la persona che indossava il prototipo esplorava le posture e i movimenti necessari a ottenere gli esiti sonori giudicati come qualitativamente migliori e di maggior stimolo per la ricerca, mentre altri partecipanti agivano in rete sul sistema per commutare le alternative di interazione sonora e modificarle ed eventualmente combinarle.

Seguiva un esperimento di design inverso: alcuni scenari di utilizzo venivano ricavati speculativamente a partire dai suoni e dai gesti esplorati fino a quel momento. Veniva selezionato uno degli scenari e studiato come raffinare le interazioni per rendere l’esperienza del giubbino potenzialmente convincente.

##### 3.1.2 Valutazioni

I partecipanti al workshop provenivano dall’ambito della musica elettroacustica e della performance con l’uso di sistemi interattivi, per la maggior parte studenti di Conservatorio. Nel corso delle attività sono emerse esperienze soggettive che meritano di essere riprese discorsivamente e di cui si dovrà tenere conto nell’organizzare un ulteriore workshop o esperimento nell’eventualità di approfondirle. A parte queste, l’iniziativa avrebbe potuto portare a risultati più robusti se si fosse svolta in un arco di tempo più esteso; infatti, i partecipanti provenienti dall’ambito artistico musicale non sono abituati a concentrare l’attività in forma laboratoriale se non in presenza di obiettivi molto

specifici e ristretti - come raffinare l'interpretazione di un brano studiato in precedenza.

Come accennato, durante il workshop veniva suggerito di utilizzare i trigger attivabili con determinati movimenti o gesti e sonificare l'azione con un suono preregistrato, poi di utilizzare i valori dei sensori come controlli parametrici di un algoritmo di sintesi per modello fisico, in una successione strumentale all'apprendimento dal semplice al complesso. A ciò è riferibile la categorizzazione discreta e continua della sonificazione considerato il tipo di algoritmo di generazione sonora e la connotazione di una specifica azione o scenario dipendente dalle varie combinazioni possibili [6].

All'interazione discreta può essere connaturata la sonificazione "discreta" allorché il suono segnala il completamento di un'azione come l'accensione di un dispositivo premendo su un interruttore, in mancanza del tipico *feedback* impulsivo di un meccanismo a molla. L'interazione continua sembra connotarsi meglio con un *feedback* sonoro "continuo" in quanto il fenomeno fisico di durata non istantanea si può denotare per una trasformazione energetica anche sonora.

In questo quadro, la possibilità di sonificare in modo continuo il gesto compiuto per un azionamento (pressione dell'interruttore) sembra conferire un'attribuzione qualitativa impropria, mentre la sonificazione discreta di un'azione continua sembra innaturale perché tende a discretizzare un gesto naturalmente continuo.

Nell'esperienza soggettiva del prototipo "scuba" la contrapposizione tra le categorie è apparsa alquanto sfumata. Seppure l'indumento include sensori azionabili in modo discreto, risultava tuttavia difficile percepire come discreta alcuna interazione con alcun tipo di sonificazione; anche il più elementare azionamento di un suono preregistrato mediante un trigger tendeva a risolvere il concetto del discreto nella continuità fisica dello svolgimento del gesto e del suono nel tempo.

Il principale fattore di ciò può essere considerato l'elevato *embodiment* dell'interazione [7]. Il compimento del singolo gesto o azione, fisiologicamente, determina il coinvolgimento dell'intero corpo. Poiché l'indumento è aderente all'intera metà superiore e non già solo a un arto o parte di esso, l'accoppiamento meccanico rende inesatta e per certi aspetti imprevedibile la correlazione tra movimento ed esito posturale e gestuale, ossia tra intenzione del soggetto e fenomeno complessivamente percepibile. Inoltre, la rigidità di alcuni sensori in special modo è bilanciata dalla cedevolezza dell'indumento e delle parti del corpo cui è accoppiata.

I valori dei sensori (per i controlli parametrici continui sugli algoritmi di elaborazione sonora) e le soglie (per i trigger) si modificano "relativamente" di continuo e il soggetto è a sua volta spinto ad adattarsi a questa scioltezza modificando nei dettagli il proprio comportamento. Per corrispondere a questa interazione probabilistica e adattiva, allo stato attuale del progetto, la soluzione di mixare tra loro più bus di controllo derivati da differenti riscaldamento dei valori dei sensori, lavorando in forma eventualmente collaborativa dalle postazioni in rete, nonostan-

te l'approccio intuitivo ed empirico, è apparsa sostenibile: ci si è riferiti quindi all'opportunità di "perdere controllo per guadagnare influenza" come discussa a proposito dell'interazione per la performance musicale [8].

Il fenomeno era apprezzabile sia dal soggetto che dagli spettatori in forza della percezione auditiva dell'interazione sonora con l'indumento, che possiede dunque un'importante valenza relazionale.

Collegata a quanto appena evidenziato è la ri-emergenza linguistica del gesto. Come i movimenti della persona che indossa l'abito possono dare luogo a gesti che trasportano significati in sostituzione della comunicazione verbale, anche mediante l'aggiunta del suono viene a crearsi un *layer* di segni che è, a ben vedere, paragonabile a quello del gesto inteso come forma di comunicazione paralinguistica in accompagnamento al linguaggio parlato.

In questa prospettiva, il gesto riacquisisce parzialmente la connotazione "discreta" che sembrava essersi risolta nella continuità fisica. L'intuizione è che la valenza paralinguistica della sonificazione mediante l'indumento aumentato vada studiata in prospettiva semiotica, confrontandosi nel particolare (ad esempio) con la letteratura che si occupa della mimica durante la conversazione.



Figura 4. Un esemplare del giubbino interattivo.

### 3.2 Presentazione

In seguito al workshop, seppure non si sia arrivati all'effettiva proposizione in pubblico, la manifestazione di interesse a presentare l'indumento nell'ambito di un festi-



**Figura 5.** Il giubbino interattivo durante l'azione musicale (performer: Lara Teković).

val del cinema di fantascienza (ottobre-novembre 2017)<sup>16</sup> ha esposto il prototipo “scuba” alla valutazione da parte degli organizzatori e permesso di comprendere le potenzialità dell’oggetto in uno dei contesti più emblematici per quanto riguarda il discorso sociale e critico sulle tecnologie indossabili.

Far apparire il prototipo durante il festival aveva lo scopo di sensibilizzare il pubblico nel merito dei risvolti culturali dell’impiego delle tecnologie indossabili in contesti *everyday* e in particolare della valenza estetico-funzionale dell’utilizzo del suono in scenari relazionali.

Ci si proponeva di mostrare il prototipo durante brevi segmenti della programmazione come un comune capo di abbigliamento indossato da una persona in capo all’organizzazione in rapporto diretto con il pubblico. L’apparizione del vestito sul palco e negli altri spazi del festival veniva considerata come un cameo; si doveva creare una narrazione implicita a corredo della quale, per chi avesse voluto approfondire, sarebbero stati messi a disposizione dei *flyer* esplicativi. Il prototipo doveva mutare il comportamento sonoro in base agli atteggiamenti posturali e alla mimica della persona che lo indossava, proponendo design sonori utili a caratterizzare in modo non invasivo lo scenario enfatizzandone gli aspetti relazionali.

Le normali tecniche di amplificazione della voce parlata avrebbero trasportato i suoni prodotti dal vestito senza il bisogno di ulteriori dispositivi. Il sistema veniva inserito in una WLAN e supervisionato da una apposita regia.

Venivano delineati due scenari:

- il vestito crea una narrazione ipertestuale composta dal sonoro di citazioni filmiche durante una tavola rotonda, *speech* o intervista (figura del moderatore);
- l’indumento viene usato come uno *smart cloth* che permette di comandare dispositivi remoti o immaginari; ai gesti mimati vengono associati suoni identi-

ficativi; i gesti vengono compiuti in stretto rapporto con l’andamento degli eventi sul palco e la regia (figura del presentatore).

### 3.3 Performance

L’opportunità di confrontare il prototipo “scuba” con le prerogative di un ambito familiare quale la musica elettroacustica e la composizione strumentale si è presentata nella cornice di un progetto artistico presso il Conservatorio di Venezia denominato “laboratorio di musica e tecnologie digitali” (24 maggio 2018).<sup>17</sup>

La cinquantesima edizione del laboratorio, dal titolo “La scala violata”, è stata concepita per lo scalone monumentale di Palazzo Pisani a Santo Stefano (sede del Conservatorio) in una forma quasi concertistica, ossia un percorso in evoluzione continua plasmato da compositori, esecutori e pubblico insieme: le opere strumentali ed elettroacustiche degli studenti del Conservatorio venivano eseguite da sedici postazioni in una scansione temporale determinata dal passaggio del pubblico e di una performer al flauto a becco mascherata che agiva al suo interno.

La peculiarità del progetto consisteva nella non congruenza di visione e ascolto, in quanto il suono era distintamente udibile lungo più segmenti dello scalone, mentre alla vista si poteva osservare l’azione degli esecutori di una postazione alla volta. La strumentazione elettroacustica enfatizzava questo aspetto.

Il giubbino faceva parte del corredo della performer e non era stato inserito nell’impianto di diffusione elettroacustica in modo da circoscrivere l’azione sonora entro i limiti fisici della diffusione del suono del flauto a becco, isolando l’interazione attorno alla figura che lo indossava.

Nello svolgimento del percorso il prototipo risultava udibile solo dalle persone più vicine. La rilevanza nel contesto

<sup>16</sup> <http://www.sciencefictionfestival.org>

<sup>17</sup> [http://www.60db.it/web\\_1/index.php?page=activity&article=workshops&lang=it](http://www.60db.it/web_1/index.php?page=activity&article=workshops&lang=it)

artistico emergeva in special modo durante la sequenza introduttiva che aveva luogo nel sottoportico del palazzo e durante le transizioni tra postazioni successive, allorché il pubblico si ridistribuiva nello spostamento e si verificavano momenti di simultaneità nell'esecuzione musicale.

L'artificio dell'indumento interattivo si integrava con facilità nel complesso del progetto artistico grazie alle capacità mimiche della performer e per la cura riservata all'*outfit*.

Il principale tipo di interazione sonora durante la performance consisteva nell'uso di trigger per azionare suoni preregistrati che dovevano enfatizzare lo straniamento della performer. Durante le prove era stata testata, con esiti soddisfacenti, l'interazione continua con il controllo parametrico dell'algoritmo di sintesi per modello fisico del suono del flauto; per ragioni di tempo, tuttavia, non si è potuto predisporre lo scenario esecutivo e si è preferito non utilizzarla.

#### 4. POSTILLA

Il massiccio coinvolgimento della ricerca scientifica nelle tecnologie indossabili e in particolare nell'area di interesse degli *electronic textile* ha prodotto un'enorme quantità di letteratura che spazia dagli aspetti più specifici delle scienze dei materiali fino alle applicazioni in prodotti e servizi negli ambiti di utilizzo più evoluti quali lo sport, la difesa militare e la medicina [9, 10, 11, 12].

Negli esempi di indumenti aumentati citati in letteratura - e ancor più nella pubblicistica divulgativa, ad es. [13, 14], si fa fatica a trovare atteggiamenti speculari e prese di posizione critiche puntuali; i miglioramenti funzionali del servizio, della qualità del prodotto o del processo di progettazione sono considerati nell'ottica di scenari futuri che si connotano per una corrispondenza finale necessariamente positiva tra tecnologie indossabili ubiqua e bisogni individuali e sociali non esaminati nel dettaglio.

Nell'ambito del design, invece, le pubblicazioni che si caratterizzano per una notevole audacia ideativa sono anche molto caute nel merito delle motivazioni e delle modalità che potranno o meno portare alla diffusione di nuove tecnologie indossabili.

Nell'editoriale di una recente edizione monografica dell'*International Journal of Design* dedicata a *wearable* e interazioni di tendenza, attestando la mancanza di un repertorio stilistico di connessioni tra uso ed estetica, si sostiene chiaramente la necessità di esplorare le tecnologie indossabili in una prospettiva primariamente socioculturale, con l'obiettivo di competere con il pensiero funzionalista e techno-futurista promosso negli ambiti di sviluppo dei materiali e delle tecnologie. Viene inoltre menzionata la totalità del corpo quale fattore su cui impostare la ricerca sulle prossime tecnologie indossabili, in opposizione all'ondata di beni di consumo che ha massimizzato il profitto basandosi su un fattore di forma - quello del gioiello - già ampiamente leggibile da parte della società [15].

Nell'ambito del progetto qui presentato, appare necessario riferirsi non solo agli strumenti sociologici e delle scienze umane, ma anche agire in base a prospettive storiche, estetiche, di teoria critica e sociale. Infatti, la società è allo stesso tempo informata su pratiche che utiliz-

zano le tecnologie indossabili secondo prospettive radicalmente diverse, differenziandosi nelle modalità di presentazione e documentazione, aventi una portata economica relativamente insignificante ma un competitivo potere di penetrazione culturale.

Un'ulteriore area di pubblicazioni, comparativamente sparsa e meno visibile nei network bibliografici, attraversa campi di ricerca potenzialmente distanti, integra tra gli oggetti di studio i fenomeni artistici dei nuovi media e la relativa pubblicistica, considera i canali di informazione non ufficiali tra cui i singolari gruppi di discussione su internet ai quali partecipano curatori e artisti di rilevanza internazionale.

Tra queste, una recente monografia [16] ha storicizzato l'area dei *wearable* con particolare riferimento agli indumenti, individuando pochi passaggi storici mediante categorie interpretative definite. Anche in questo studio appaiono fondamentali la distinzione tra funzionalità ed estetica e l'adozione strumentale delle due da parte di ambiti spesso opposti - ad esempio, il terreno della moda e specialmente le sfilate vengono considerati quali complesso e non facilmente gestibile strumento per divulgare e ottenere consenso. Il testo rende conto dell'interpretazione delle tecnologie indossabili come strumento di critica sociale ed economica nell'ambito del grande numero di progetti non perfettamente posizionabili nell'intervallo che va dall'opera d'arte, all'etica DIY e DIWO, all'attivismo politico-culturale.

Poiché molte di queste esperienze non hanno dispiegato il loro potenziale, o perché non sono realizzate compiutamente o perché scarsamente documentate, vale la pena menzionare una serie di chiamate ai progetti "fallimentari", bandita in tempi recenti per porli all'attenzione generale nell'ambito di esposizioni permanenti all'intersezione tra festival e conferenza.<sup>18</sup>

Considerate le prerogative estetiche e poetiche dell'ambito in cui si svolge, appare evidente che il posizionamento costituisca una problematica da affrontare anche per il progetto qui descritto. Riguardo all'introduzione dell'argomento del Sonic Interaction Design nel contesto di un'academia di musica, se essa ha potuto stimolare l'innovazione del percorso artistico, è anche vero che il contributo degli artisti, all'interno di un processo di progettazione partecipativo, costituisce un aspetto primario ancora da approfondire. Inoltre, vista la varietà dei possibili esempi di applicazione del prototipo "scuba" e la necessità di attraversare obliquamente il campo artistico e la ricerca industriale, per rendere i futuri passaggi di questo progetto elementi utili a una ricerca autenticamente scientifica sarà opportuno riferirsi all'ambito disciplinare delle scienze del design incorporandone alcune metodologie; diversamente, il progetto dovrà enfatizzare la propria valenza estetica e critica, posizionandosi più nettamente nell'ambito delle emergenti pratiche di ricerca artistica [17].

<sup>18</sup> <http://trials-and-errors.com/concept>. [...] *Our approach reverses the common R&D constructivist methods, by using deconstruction as a tool for critical inquiry in the field of wearable technologies and e-textiles. [...] The project uses art's capacity to critically and, at the same time, poetically and self-ironically, address contemporary challenges and concerns.*

## 5. BIBLIOGRAFIA

- [1] G. Klauer, A. Metus, and P. Polotti, "Sonic interaction design e paper wearables," in *Extending interactivity. Atti del XXI CIM - Colloquio di Informatica Musicale* (A. Terzaroli and A. Valle, eds.), pp. 33–40, AIMI - Associazione Informatica Musicale Italiana, DADI - Dip. Arti e Design Industriale. Universit IUAV di Venezia, 2016.
- [2] G. Klauer, A. Metus, and P. Polotti, "Sonic interaction design for paper wearables," in *Proceedings of the 12th International Audio Mostly Conference on Augmented and Participatory Sound and Music Experiences*, pp. 49:1–49:7, ACM, 2017.
- [3] L. E. Dunne, C. Simon, and G. Gioberto, "E-textiles in the apparel factory: Leveraging cut-and-sew technology toward the next generation of smart garments," in *Fundamentals of Wearable Computers and Augmented Reality, Second Edition* (W. Barfield, ed.), pp. 619–638, CRC Press, 2015.
- [4] S. Baldan, S. Delle Monache, and D. Rocchesso, "The sound design toolkit," *SoftwareX*, vol. 6, pp. 255–260, 2017.
- [5] P. Cook, "A meta-wind-instrument physical model, and a meta-controller for real time performance control," in *Proceedings of the 1992 International Computer Music Conference*, pp. 273–276, 1992.
- [6] D. Rocchesso, P. Polotti, and S. Delle Monache, "Designing continuous sonic interaction," *International Journal of Design*, vol. 3, no. 3, 2009.
- [7] P. Dourish, *Where The Action Is: The Foundations of Embodied Interaction*. Cambridge (MS): MIT Press, 2001.
- [8] A. de Campo, "Lose control, gain influence - concepts for metacontrol," in *Music Technology Meets Philosophy: From digital echos to virtual ethos. Proceedings of the 40th International Computer Music Conference joint with the 11th Sound and Music Computing conference* (A. Georgaki and G. Kouroupetroglou, eds.), pp. 217–222, 2014.
- [9] T. Dias, ed., *Electronic Textiles: Smart Fabrics and Wearable Technology*. Woodhead Publishing, 2015.
- [10] J. McCann and D. Bryson, eds., *Smart clothes and wearable technology*. Woodhead Publishing, 2009.
- [11] D. James and N. Petrone, *Sensors and Wearable Technologies in Sport: Technologies, Trends and Approaches for Implementation*. Springer, 2016.
- [12] S. C. Mukhopadhyay, ed., *Wearable Electronics Sensors For Safe and Healthy Living*. Springer, 2015.
- [13] D. Guler, M. Gannon, and K. Sicchio, *Crafting Wearables Blending Technology with Fashion*. Apress, 2016.
- [14] K. Hartman, *Make: Wearable Electronics*. Maker Media, 2014.
- [15] O. Tomico, L. Hallnäs, R.-H. Liang, and S. A. Wensveen, "Towards a next wave of wearable and fashionable interactions," *International Journal of Design*, vol. 11, no. 3, 2017.
- [16] S. E. Ryan, *Garments of Paradise. Wearable discourse in the digital age*. Cambridge (MS): MIT Press, 2013.
- [17] L. Candy and E. Edmonds, "Practice-based research in the creative arts. foundations and futures from the front line," *Leonardo*, vol. 51, no. 1, pp. 63–69, 2018.

# TECNOLOGIE PER LA DIDATTICA MUSICALE: UN'ESPERIENZA CON LA REALTÀ VIRTUALE

**Michele Geronazzo**

**Rolf Nordhal**

**Stefania Serafin**

Dept. of Architecture, Design & Media Tech.

Aalborg University Copenhagen

{mge, rn, sts}@create.aau.dk

**Edoardo Degli Innocenti**

**Diego Vescovi**

Dept. of Information Eng.

University of Padova

edosigma@gmail.com

ves.diego@gmail.com

**Federico Avanzini**

Dept. of Computer Science

University of Milano

avanzini@di.unimi.it

## ABSTRACT

La mobile virtual reality (VR) sta diventando sempre più popolare e accessibile a tutti coloro che possiedono un nuovo smartphone. In particolare, la didattica digitale può trarre vantaggio dall'interazione naturale e dall'immersione in ambienti virtuali, a partire dalla scuola primaria. Questo articolo presenta un quadro tecnico e metodologico a supporto dell'uso della mobile virtual reality nell'istruzione primaria, volto a migliorare l'apprendimento musicale in termini di ascolto attivo, attenzione e tempo. Viene innanzitutto presentata un'applicazione multiplatforma chiamata VR4EDU, di cui si descrivono le scelte di progetto e lo sviluppo delle applicazioni di virtual reality, con particolare attenzione a una innovativa soluzione di "locomozione virtuale" che risulta essere intuitiva e usabile. Le applicazioni di VR4EDU possono funzionare su hardware a basso costo, diventando quindi appetibili per un uso didattico. L'approccio proposto è stato validato tramite uno studio con due classi dell'ultimo anno di scuola primaria, incentrato sull'identificazione e l'apprendimento dei generi musicali. Una valutazione preliminare dei risultati mostra un miglioramento nell'apprendimento della caratterizzazione del genere rispetto alle lezioni tradizionali basate su materiale stampato e ascolto passivo.

## 1. INTRODUZIONE

In molti paesi occidentali, l'educazione musicale ha subito tagli dei finanziamenti pubblici, seguendo una tendenza generale a deprivilegiare la musica e le altre attività legate all'arte [1]. Questi tagli sono negativi, soprattutto considerati i molti potenziali benefici dell'educazione musicale per i bambini. Ricerche precedenti mostrano che l'educazione musicale è utile per migliorare le capacità cognitive spazio-temporali [2], anche se i benefici nel migliorare le prestazioni in altre materie, come ad esempio la matematica, non sono stati ancora dimostrati. L'educazione musicale comprende sia l'apprendimento di uno strumento che imparare a conoscere la teoria della musica e la sua storia.

*Copyright: ©2018 Michele Geronazzo et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution License 3.0 Unported](https://creativecommons.org/licenses/by/3.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.*

La Realtà Virtuale (VR) è stata recentemente adottata per insegnare diverse materie, che vanno dalla storia dell'arte [3] all'anatomia [4] e ad altre discipline scientifiche che richiedono l'uso di un laboratorio [5].<sup>1</sup> La musica, tuttavia, non rientra tra i sopracitati argomenti. La principale domanda di ricerca di questo articolo è se la VR può essere efficacemente sfruttata per l'apprendimento della musica nelle scuole primarie, con un focus sulla teoria e la storia della musica. In particolare proponiamo una nuova applicazione di VR, denominata VR4EDU di seguito, che viene utilizzata in questo studio per facilitare l'apprendimento dei generi musicali.

La maggior parte delle applicazioni educative sviluppate finora usa dispositivi di VR piuttosto costosi. Infatti una delle ragioni per cui le tecnologie VR sono rimaste finora lontane dalla portata delle scuole primarie è la fattibilità finanziaria [6]. Tuttavia, questa limitazione è ora in fase di superamento grazie allo sviluppo di tecnologie hardware a basso costo e alla disponibilità di pacchetti software per la simulazione di ambienti 3D. Ad esempio, per le attività in classe sono state pensate appositamente soluzioni come i visori VR di cartone (che possono essere messi a disposizione di un'aula) in combinazione con smartphone (tipicamente già posseduti dagli alunni o dalle loro famiglie), facilitando esperienze interattive e personalizzate di apprendimento immersivo [7, 8]. La tecnologia di VR per smartphone, in particolare, può essere facilmente introdotta nelle dinamiche di insegnamento in classe [9]. In questo lavoro seguiamo questo approccio e utilizziamo visori VR mobili disponibili in commercio in combinazione con software multiplatforma gratuiti.

Ci rivolgiamo a parte del programma nazionale di musica per le scuole primarie italiane, con specifico riferimento a un sottoinsieme di obiettivi di apprendimento, conoscenze, abilità e attività, evidenziati in corsivo nella Tabella 1.

Il punto (1) riguarda la conoscenza degli strumenti musicali: VR4EDU approfondisce la conoscenza di strumenti legati a specifici generi musicali. Il punto (2) si riferisce all'esperienza di ascolto, tradizionalmente basata sulla riproduzione mono o stereo in classe: con VR4EDU l'esperienza di ascolto è individuale, in quanto la musica viene trasmessa attraverso auricolari e arricchita da un ambiente immersivo. Il punto (3) ha a che fare con la cultura musicale, in particolare con i generi musicali, ed è l'obiettivo

<sup>1</sup> Una risorsa preziosa è <https://vrschoolresearch.com/>

**Tabella 1:** Porzione del programma di musica della scuola pubblica. Le voci numerate e in grassetto rappresentano parti specifiche del programma prese in considerazione da VR4EDU.

| <i>Obiettivi di apprendimento</i>   | <i>Conoscenza</i>   | <i>Competenze</i>   | <i>Attività</i>   |
|---|---|---|---|
| Essere in grado di cantare conoscendo le regole della musica.                   | <b>(1) Conoscere le caratteristiche sonore e le famiglie di strumenti.</b>  | Saper rispettare le regole della direzione musicale   | <b>(2) Ascolto, registrazione dati, registrazione di programmi specifici con il computer, il LIM o il CD.</b> |
| Essere in grado di seguire correttamente il ritmo.                              | Conoscere la notazione convenzionale.   | <b>(1) Sapere riconoscere l'appartenza di uno strumento alla sua famiglia.</b>  | Canzoni da esercitazione eseguite con strumentazione d'aula, con flauto o con strumenti personali.            |
| Essere in grado di suonare strumenti ritmici facili o altri strumenti melodici. | Conoscere le componenti antropologiche della musica.<br><b>(3) Conoscere canzoni appartenenti a un repertorio variegato di genere popolare.</b> | Sapere utilizzare risorse espressive nell'intonazione di canti semplici sia singolarmente che in gruppo, nel rispetto delle regole stabilite. | Esercitazione di canto corale o individuale con l'uso del microfono.  |
|   | Conoscere le quattro dimensioni del suono (intonazione, timbro, volume, durata).  | Sapere leggere e interpretare testi di brani legati ad uno specifico contesto storico-sociale della canzone.                                  | <b>(4) Attività grafica-pittorica, attività di danza o interpretazione con il corpo.</b>                      |
|   |   | Sapere eseguire brani semplici con il flauto, da solo o in gruppo.  |   |

principale dell'applicazione: VR4EDU espone gli alunni a brani musicali che appartengono a generi diversi e possono essere ascoltati ed esplorati in modo interattivo. Infine, il punto (4) riguarda le attività di apprendimento, in particolare le attività di movimento con la musica: VR4EDU utilizza i movimenti del corpo per consentire agli alunni di muoversi all'interno dell'ambiente virtuale 3D, incoraggiandoli a ballare sul posto e spostarsi mentre ascoltano la traccia di musica di sottofondo.

## 2. LAVORI CORRELATI

Il potenziale delle nuove tecnologie per migliorare i risultati e l'apprendimento degli studenti – se usate in modo appropriato – è stato riconosciuto da molto tempo [10]. In particolare, la sperimentazione della tecnologia VR nell'istruzione primaria, secondaria e superiore è iniziata nei primi anni '90 [11, 12], con l'utilizzo di visori, guanti e tute "full-body". Nel complesso la ricerca precedente mostra alcuni risultati incoraggianti per quanto riguarda l'uso della VR in contesti educativi [13]. Una meta-analisi sull'efficacia dell'uso della VR nell'insegnamento [14] ne conferma il potenziale, ancor di più quando si utilizzano ambienti basati su giochi piuttosto che simulazioni.

L'immersione offerta dalle tecnologie VR offre forti benefici che possono supportare gli studenti [7]: questa promuove l'apprendimento attraverso un maggiore coinvolgimento e facilita il trasferimento di conoscenze dalla classe al mondo reale; inoltre promuove esperienze educative che attingono all'*apprendimento in situ* e all'*apprendimento attraverso il fare*. La prima definizione si riferisce all'apprendimento che avviene in una comunità di pratica, nello stesso contesto in cui viene applicata [15], mentre la seconda postula che il coinvolgimento attivo degli studenti nella costruzione di nuove conoscenze li aiuta a padroneggiare, mantenere e generalizzare tali conoscenze [16]. Questi approcci contrastano con gli ambienti didattici tradizionali in cui gli studenti imparano per assimilazione, cioè ascoltando passivamente un istruttore.

Riguardo ad applicazioni di VR per la musica in particolare, la maggior parte sono mirate a suonare strumenti musicali virtuali piuttosto che allo studio della teoria della musica. Ad esempio, Music Room<sup>2</sup> è una raccolta di strumenti percussivi con cui interagire utilizzando i "controller" dello smartphone. Inoltre, la filosofia dell'applicazione è che agisce come un MIDI controller per l'utilizzo in qualsiasi Workstation Audio Digitale (DAW). Un'applicazione simile è Soundstage VR,<sup>3</sup> che, oltre agli strumenti VR interattivi, include anche una catena di mix modulari con una libreria di effetti ed elaborazioni, nonché una fase di looping e registrazione da utilizzare in post-produzione o in altre produzioni multimediali.

Queste applicazioni promuovono il gioco e l'interazione con strumenti musicali VR per produrre musica o per allenare abilità ritmiche, con uno scopo commerciale. D'altra parte, la ricerca accademica ha sviluppato nel corso degli anni diversi strumenti musicali VR, con lo scopo principale di creare simulazioni o estensioni di quelli esistenti [17]. Inoltre, in passato sono stati proposti anche ambienti virtuali per imparare a suonare strumenti tradizionali, ad esempio un flauto [18] e un tutor per pianoforte [19], per citarne solo alcuni.

La realtà aumentata, dove uno strumento tradizionale è "arricchito" con la tecnologia, sembra essere più adatta della realtà virtuale per insegnare a suonare uno strumento musicale. Ciò è dovuto principalmente al fatto che gli strumenti musicali sono interfacce estremamente sofisticate che si sono sviluppate nel corso dei secoli, mentre le interfacce VR sono in confronto ancora primitive. Per questo in letteratura sono apparsi diversi sistemi, come Andante e Andantino [20, 21], dove la realtà aumentata viene utilizzata per proiettare personaggi animati su una tastiera per pianoforte, al fine di coinvolgere i bambini non solo nel suonare la giusta nota, ma anche nelle caratteristiche immancabili di una performance musicale.

Naturalmente l'immersività è fondamentale anche per il canale uditivo. La letteratura adotta il termine "auralizza-

<sup>2</sup> [musicroomvr.com/](http://musicroomvr.com/)

<sup>3</sup> [soundstagevr.com/](http://soundstagevr.com/)

zione” per riferirsi al rendering audio spaziale in ambienti virtuali [22]. Le tecnologie audio spaziali permettono la direzione dell’attenzione degli utenti e il miglioramento del realismo dell’esperienza VR con un’influenza positiva sul carico di lavoro, le prestazioni e la presenza [23]. Inoltre, l’interazione circolare tra presenza ed emozioni è ben nota nella letteratura scientifica e porta a considerare la realtà virtuale come un mezzo affettivo [24], capace cioè di interagire con gli stati affettivi [25] e i processi di memoria [26]. Nel caso del nostro studio in particolare, poiché la musica evoca forti emozioni, può anche essere coinvolta nella formazione di ricordi su informazioni associate a un particolare contesto, genere musicale o brano [27]. Per tutte queste ragioni, le interazioni affettive tra memoria contestuale e contenuti musicali rendono la navigazione attiva in VR uno strumento efficace per l’apprendimento della musica.

### 3. L’APPLICAZIONE VR

Abbiamo sviluppato un’esperienza di ascolto VR coinvolgente, arricchita da modelli 3D visivamente accattivanti di strumenti musicali. Abbiamo progettato due diverse applicazioni VR, “Musical Labyrinth Exploration” (MLE in seguito) e “Room Learning Experience” (RLE in seguito). La prima rappresenta un’esperienza preliminare in cui gli studenti possono imparare a muoversi nell’ambiente 3D di VR4EDU, mentre la seconda è l’effettiva applicazione di apprendimento dei generi musicali, da usare come integrazione al programma scolastico tradizionale.

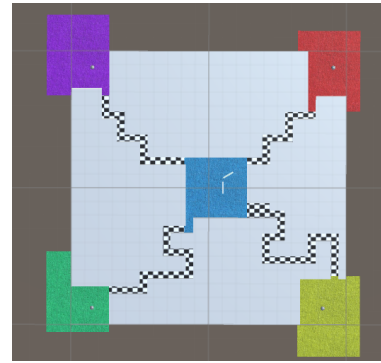
Questa sezione presenta gli elementi principali e gli aspetti innovativi delle applicazioni.

#### 3.1 Scenari VR e interazioni

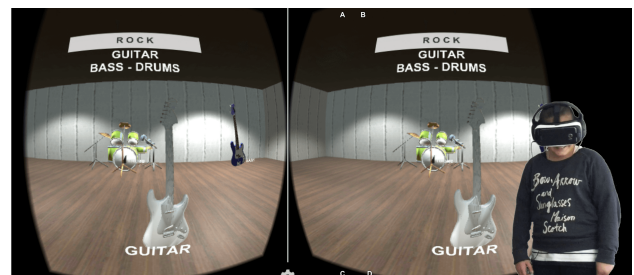
Lo scenario MLE è un labirinto 3D che aiuta ad apprendere il sistema di navigazione, in modo che gli utenti abbiano l’opportunità di prendere confidenza con l’esplorazione dell’ambiente 3D. A questo scopo, l’applicazione propone il seguente compito: all’utente viene chiesto di trovare quattro diverse aree colorate in cui sono posizionate quattro sfere sospese a mezz’aria, che rappresentano le sorgenti sonore di quattro diversi brani musicali (si veda la Fig. 1 per una rappresentazione schematica della mappa).

Questo compito è pensato per incoraggiare l’utente a esplorare l’intero labirinto, imparando progressivamente a evitare i muri e a compiere percorsi stretti e articolati. Per rendere facilmente distinguibili i brani musicali, abbiamo selezionato quattro timbri diversi: pianoforte, chitarra classica, violino, flauto. Per incoraggiare ulteriormente l’esperienza di esplorazione, le tracce vengono rese udibili attraverso le pareti in modo che l’utente possa comprendere la direzione del suono in ingresso anche quando la sfera galleggiante non è visibile. Di conseguenza, gli utenti sono istruiti a prestare attenzione sia allo stimolo visivo che a quello uditivo per la navigazione, sfruttando la sorgente sonora come un segnale acustico di aiuto al raggiungimento dell’obiettivo [28].

Lo scenario RLE è composto da uno spazio rettangolare limitato dove l’utente può ascoltare un brano rappresenta-



**Figura 1:** Vista dall’alto del labirinto MLE. Le caselle rossa, viola, verde e gialla sono le quattro camere e la casella azzurra è la stanza di partenza situata al centro. Il percorso a labirinto è raffigurato con il pavimento a scacchiera. In ogni stanza è visibile un piccolo punto che rappresenta la sfera della sorgente audio.



**Figura 2:** Split screen dall’applicazione RLE. Qui è visibile il titolo principale con il nome del genere e i suoi principali strumenti caratterizzanti.

tivo di uno specifico genere musicale. Inoltre, i modelli 3D degli strumenti musicali appartenenti ad un genere specifico vengono posizionati nello spazio come se fossero collocati su un palcoscenico, nella consueta disposizione di una tipica performance reale. Abbiamo deciso di limitare lo spazio esplorabile in modo da aiutare l’utente a concentrarsi sugli strumenti e sul loro suono, riducendo al minimo le distrazioni.

All’utente viene inizialmente dato un messaggio di benvenuto che introduce la stanza e un messaggio sempre acceso con il titolo del genere musicale di quella stanza, insieme ai suoi principali strumenti caratterizzanti. Queste informazioni testuali possono essere lette sempre in un colpo d’occhio, inoltre vicino ad ogni modello 3D viene posizionato il nome corretto dello strumento corrispondente (si veda la Fig. 2 per uno esempio).

L’applicazione RLE propone otto diverse “stanze di genere”, insieme ai rispettivi strumenti tipici del genere in questione. Lo spazio architettonico di ogni stanza non varia da un genere all’altro. Cambiano solo titolo, etichette e strumenti musicali (con la loro disposizione spaziale).

#### 3.2 Implementazione

L’hardware usato include tre smartphone: un iPhone 6s, un LG G2 e un Samsung Galaxy S7 Edge, tutti equipaggiati con un display ad alta risoluzione e in grado di reg-

gere il carico computazionale dell'applicazione. Il feedback uditivo è stato reso con cuffie circumaurali Sennheiser HD 429, Beats by Dr. Dre e Satechi Wireless, connesse direttamente agli smartphone.

Lo sviluppo software è basato su Blender 2.77,<sup>4</sup> programma di grafica 3D open source molto diffuso, e Unity 5.34f1<sup>5</sup> con la SDK per Google Cardboard.<sup>6</sup> Blender è stato usato per creare i principali modelli 3D (pareti del labirinto e delle stanze, ecc.). Unity ha il vantaggio di essere cross-piattaforma e ha permesso di compilare VR4EDU sia per Apple iOS che per Android OS.

Per quanto riguarda il rendering audio, le quattro sorgenti sonore associate alle sfere nell'applicazione MLE sono state realizzate utilizzando le funzionalità native di Unity. In particolare abbiamo usato una legge logaritmica per l'intensità come funzione della distanza, con il parametro "max-distance" (la distanza a cui l'intensità smette di decrescere) regolato a 50 m e corrispondente a intensità nulla. Quindi non si sente alcun suono oltre i 50 m di distanza, e questo assicura anche che le quattro tracce audio non possano essere udite tutte contemporaneamente.

Per l'applicazione RLE abbiamo usato regolazioni diverse: ancora una legge logaritmica di decadimento dell'intensità, ma questa volta con i parametri "min-distance" e "max-distance" regolati rispettivamente a 30 m e 50 m. Inoltre l'intensità oltre i 50 m è regolata a un valore non nullo, cosicché è sempre possibile sentire la traccia anche a distanze elevate. Questo invita l'utente ad avvicinarsi alla sorgente sonora. Per distanze minori di 30 m, invece, l'intensità è costante al livello massimo. Le sorgenti sonore sono posizionate al centro della stanza, in particolare nel baricentro dei modelli 3D degli strumenti.

Al fine di ottimizzare il carico computazionale, alcuni modelli 3D sono stati semplificati decimando vertici e facce (usando Blender). Questo si è reso necessario dal momento che test iniziali hanno evidenziato che modelli 3D troppo dettagliati diminuivano il frame-rate della visualizzazione deteriorando l'esperienza di navigazione. Abbiamo validato la nostra procedura di decimazione attraverso un test in cui abbiamo misurato il frame-rate in una simulazione di navigazione all'interno della Stanza Rock di RLE, usando lo smartphone più performante e quello meno potente, ovvero Samsung (1440 × 2560 pixel, circa 534 ppi, processore Samsung Exynos 8 Octa 8890 Octa Core 2.3 GHz 64bit, Mali T880 MP12) e LG (1080 × 1920 pixel, circa 424 ppi, processore Quad-core 2.26 GHz Krait 400, GPU Adreno 330). Il frame-rate massimo (quando nessun modello 3D deve essere renderizzato) è stato impostato a 60 fps. I risultati mostrano che usando i modelli decimati il frame-rate non scende mai sotto i 59 fps su entrambi gli smartphone, mentre senza decimazione scende a valori molto bassi (3-4 fps).

### 3.3 Navigazione

Uno degli aspetti chiave dell'applicazione VR4EDU è la progettazione di un sistema di navigazione affidabile e in-

tuitivo in grado di fornire indicazioni di orientamento spaziale per supportare l'usabilità e l'apprendimento [29]. L'adozione di un visore VR richiede il design di un'efficace "locomozione virtuale", in quanto il modo in cui le persone si muovono nell'ambiente virtuale è fondamentale per migliorare l'esperienza di apprendimento e prevenire problemi come la "motion sickness" e la perdita di equilibrio.

Il problema della locomozione virtuale è stato ampiamente studiato [30]. Abbiamo considerato diverse opzioni, in particolare prendendo ispirazione da interfacce comuni come telecomandi e joystick, così come approcci che prevedono di camminare sul posto [31]. La maggior parte di queste soluzioni richiede un dispositivo di input, che tiene costantemente occupate le mani degli utenti. Per offrire la più naturale e intuitiva esperienza di movimento per la navigazione e l'interazione, pur limitando i movimenti nel mondo reale [32], abbiamo progettato una soluzione che sfrutta i sensori inerziali dello smartphone all'interno del visore VR e una serie di semplici gesti del corpo. In particolare, abbiamo definito i seguenti gesti:

- Inclinare la testa in avanti/indietro → movimento rettilineo avanti/indietro.
- Ruotare la testa a destra/sinistra sul piano coronale → movimento rettilineo a destra/sinistra.
- Voltare la testa a destra/sinistra → rotazione della vista a destra/sinistra.

Al fine di rendere questa soluzione utilizzabile e prevenire instabilità, abbiamo fatto sì che non si verifichi alcun movimento tangenziale se l'inclinazione è inferiore a una soglia prefissata.

Più formalmente: si consideri un sistema di riferimento solidale con la testa dell'utente, in cui  $X$  è l'asse destra-sinistra,  $Y$  è l'asse alto-basso e  $Z$  è l'asse fronte-retro. Definiamo con  $a_{x,y,z}$  le accelerazioni misurate dai sensori inerziali lungo questi assi (queste assumono valori nell'intervallo  $[-1, +1]$  e rappresentano frazioni dell'accelerazione di gravità<sup>7</sup>). Dette  $v_{x,y,z}$  e  $\Delta_{x,y,z}$  rispettivamente le velocità tangenziali di spostamento nell'ambiente virtuale e le soglie di inclinazione, le equazioni di movimento vengono scritte come segue:

$$v_x(t) = \begin{cases} (a_x(t) \mp \Delta_x) \cdot b^{-a_x(t)} & |a_x(t)| > \Delta_x \\ 0 & \text{altrove,} \end{cases} \quad (1a)$$

$$v_y(t) = 0 \quad \text{sempre,} \quad (1b)$$

$$v_z(t) = \begin{cases} (a_z(t) \mp \Delta_z) \cdot b^{-a_z(t)} & |a_z(t)| > \Delta_z \\ 0 & \text{altrove.} \end{cases} \quad (1c)$$

Sperimentalmente abbiamo rilevato che per movimenti tipici  $a_z$  non eccede mai il valore 0.53. Abbiamo poi usato i valori  $\Delta_z = \Delta_x = 0.1$  e  $b = 2$ , che producono una velocità tangenziale massima di circa 25 m/s.<sup>8</sup>

Questo particolare valore di velocità massima è stato scelto sulla base di un test di usabilità con un gruppo di cinque bambini di 10-11 anni, che non hanno partecipato alla valutazione successiva. Il test ha proposto tre stanze

<sup>4</sup> www.blender.org/foundation

<sup>5</sup> unity3d.com

<sup>6</sup> developers.google.com/vr/unity

<sup>7</sup> docs.unity3d.com/Manual/MobileInput.html

<sup>8</sup> docs.unity3d.com/ScriptReference/CharacterController.Move.html

| Genere        | Brano                            | Strumenti                         |
|---------------|----------------------------------|-----------------------------------|
| Blues         | Sweet home Chicago               | “Cigar Box”, Armonica             |
| Trio classico | Haydn trio n. 39                 | Clavicembalo, Violino, Viola      |
| Country       | Fly around my pretty little miss | Banjo, Violino                    |
| Disco         | Togheter forever                 | Sintetizzatore                    |
| Folk          | Cincirinella teneva teneva       | Tamburello, Fisarmonica           |
| Jazz          | Take five                        | Sassofono, Contrabbasso           |
| Rock          | Satisfaction                     | Basso el., Chitarra el., Batteria |
| Swing         | Singing on nothing               | Tromba                            |

**Tabella 2:** Generi musicali considerati, brani proposti in aula e strumenti musicali caratterizzanti.

quadrate di diverse misure e diverse velocità di navigazione. I risultati hanno mostrato una preferenza una stanza di dimensioni 50 × 50 m esplorata a velocità massima di 25 m/s. In particolare questo significa che la stanza può essere percorsa in 2 s, e questo valore corrisponde al massimo tempo di reazione a uno stimolo visivo per bambini di questa età [33].

#### 4. UN ESPERIMENTO IN AULA

L'applicazione VR4EDU è stata preliminarmente collaudata in una scuola elementare, con l'obiettivo di confrontare i risultati di apprendimento di una classe esposta a didattica tradizionale con quelli di una seconda classe esposta all'uso dell'applicazione. Come caso di studio abbiamo scelto l'apprendimento di generi musicali.

##### 4.1 Protocollo sperimentale

Complessivamente sono stati coinvolti 36 bambini (18 per classe) di quinta elementare, tutti di età compresa tra i 10 e gli 11 anni assieme ai loro insegnanti. La prima classe (il “gruppo di controllo”) ha svolto le lezioni usando esclusivamente metodi tradizionali, mentre la seconda (il “gruppo VR”) ha fatto uso di VR4EDU come strumento integrativo alle lezioni tradizionali. Entrambe le classi comprendevano alunni con Bisogni Educativi Speciali (BES): 4 la prima, 6 la seconda.

Il protocollo didattico è stato articolato in quattro lezioni per ciascuna delle due classi, seguite dallo stesso insegnante (con l'aggiunta di due tecnici di supporto per il gruppo VR). La Tabella 2 riassume i generi musicali considerati, i corrispondenti brani musicali proposti agli alunni e gli strumenti musicali caratterizzanti.

Durante la prima lezione entrambi i gruppi hanno svolto un pre-test (descritto più avanti) finalizzato a quantificare le loro conoscenze di partenza. In aggiunta il gruppo VR ha svolto una sessione di ambientamento con la realtà virtuale, tramite l'applicazione MLE.

Nella seconda lezione entrambi i gruppi hanno analizzato i generi Blues, Trio classico, Country, Rock. In particolare ne hanno analizzato la storia, gli strumenti caratterizzanti e la loro disposizione tipica sul palco. A ogni genere sono stati dedicati circa 15 minuti. Gli alunni del

gruppo di controllo hanno ascoltato i corrispondenti brani attraverso un diffusore, mentre quelli del gruppo VR hanno usato a turno l'applicazione RLE per ascoltare i brani (circa 2 minuti per genere). Ad esempio, nell'analisi del Blues gli alunni hanno esplorato la “stanza del Blues” all'interno dell'applicazione RLE. Al termine degli ascolti, a entrambi i gruppi è stato chiesto di ritagliare sagome di strumenti musicali caratterizzanti ciascun genere e di incollarle seguendone la tipica disposizione sul palco.

La terza lezione si è svolta con le medesime modalità della seconda. In questo caso entrambi i gruppi hanno analizzato i generi Disco, Jazz, Folk, Swing.

Infine, nella quarta lezione entrambi i gruppi hanno svolto un post-test (descritto più avanti) finalizzato a quantificare le conoscenze acquisite. Il post-test aveva la stessa struttura del pre-test iniziale, seppure con domande diverse, e ha permesso di valutare eventuali differenze di apprendimento tra i due gruppi. In aggiunta entrambi i gruppi hanno risposto a un questionario (pure descritto più avanti), finalizzato a raccogliere dati qualitativi sull'esperienza di apprendimento.

##### 4.2 Test e questionario

Il modello di test (pre- e post-) proposto agli alunni è quello normalmente usato dall'insegnante per valutare le conoscenze degli alunni. Si articola in tre esercizi.

1. **Identificazione** del genere: gli alunni ascoltano quattro brani e, al termine di ognuno, scrivono il genere corrispondente.
2. **Associazione** tra strumenti e genere: gli alunni collegano graficamente le immagini di sette strumenti musicali ai generi in cui questi sono usati.
3. **Disposizione spaziale** degli strumenti sul palco: gli alunni devono indicare la posizione (primo piano o secondo piano) di otto strumenti associati ai vari generi (ad esempio, chitarra elettrica, basso elettrico e batteria nel Rock).

L'unica differenza tra pre-test e post-test è stata nella scelta di generi e strumenti. Questa differenziazione è stata introdotta per non rendere banale l'esecuzione del post-test.

Il punteggio del test è stato calcolato come semplice percentuale delle risposte corrette, su un totale di 19 risposte (4, 7 e 8 risposte per il primo, il secondo e il terzo esercizio). Abbiamo valutato anche la percentuale di risposte corrette per ciascun singolo esercizio.

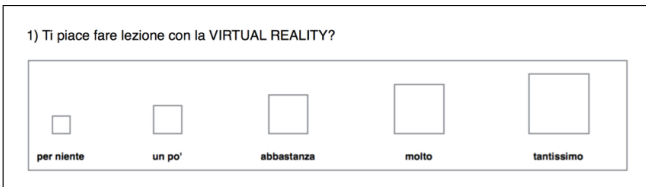
In aggiunta ai test quantitativi, l'esperienza di apprendimento è stata valutata qualitativamente tramite un questionario composto da 7 domande proposte a entrambi i gruppi e due domande aggiuntive proposte solo al gruppo VR. Le prime 4 domande (D1-4) richiedevano una risposta su una scala di Likert a 5 punti, mentre le successive 3 (D5-7) erano domande aperte.

D1: Quanto ti è sembrato difficile imparare i diversi generi musicali?

D2: Durante le attività quanto ti sei impegnato?

D3: Vorresti scoprire altri generi musicali?

D4: Quanto ti sei divertito durante le lezioni di musica sui generi musicali?



**Figura 3:** Presentazione grafica delle risposte su una scala di Likert a 5 punti.

D5: Scrivi le cose che ti sono piaciute di più durante queste lezioni sugli strumenti musicali.

D6: Scrivi le cose che non ti sono piaciute, o cosa cambieresti in queste lezioni di musica.

D7: Conosci la Virtual Reality? Cosa ne pensi?

La Fig. 3 mostra la presentazione grafica fornita agli alunni per rispondere alle prime 4 domande su una scala di Likert a 5 punti. La dimensione crescente dei riquadri aiuta il bambino a meglio comprendere le diverse risposte [34].

Infine, le due domande aggiuntive proposte solo al gruppo VR hanno utilizzato la stessa scala di Likert e sono state le seguenti:

D8: Ti piace fare lezione con la Virtual Reality?

D9: Secondo te la Virtual Reality è utile per imparare i diversi generi musicali e gli strumenti?

### 4.3 Risultati preliminari

I risultati del pre-test e del post-test sono riassunti in Fig. 4.

In particolare la Fig. 4(a) mostra i punteggi globali ottenuti nei due test da ciascun alunno del gruppo di controllo e dal gruppo VR. Si può immediatamente notare che gli alunni del gruppo VR hanno incrementato in maniera significativa i propri punteggi tra il pre- e il post-test, senza alcuna eccezione. In media, la percentuale di risposte corrette è stata circa 42% nel pre-test, e 65% nel post-test. Questo risultato suggerisce che il gruppo VR abbia appreso in maniera efficace i contenuti delle lezioni.

In contrasto, gli alunni del gruppo di controllo hanno mostrato risultati meno positivi, con un incremento quasi nullo dei punteggi tra il pre- e il post-test. In media, la percentuale di risposte corrette è stata circa 48% sia nel pre-test che nel post-test. Anche questo risultato contribuisce a mostrare che le lezioni integrate dall'uso di VR4EDU sono state più efficaci delle lezioni tradizionali. Tale conclusione è ulteriormente corroborata dalla circostanza che il gruppo VR partiva da un livello di conoscenza generalmente inferiore rispetto al gruppo di controllo, come mostrato dai risultati del pre-test.

Le Fig. 4(b,c,d) mostrano il dettaglio dei risultati per ciascuno dei tre esercizi che componevano il test. In tutti il gruppo VR ha avuto risultati migliori del gruppo di controllo. In particolare l'Esercizio 1 è stato il più impegnativo, al punto che gli alunni del gruppo di controllo hanno peggiorato il proprio punteggio tra il pre- e il post-test. Al contrario l'Esercizio 3 è risultato essere quello più facile.

Una prima analisi del questionario mostra alcune differenze tra i due gruppi. Le domande D1-3 hanno raccolto punteggi simili tra i due gruppi, con una tendenza a un maggior coinvolgimento e una minore percezione di diffi-

coltà da parte del gruppo VR. La domanda D4 ha invece mostrato una netta differenza tra i punteggi, a indicare un livello di divertimento marcatamente più alto per il gruppo VR. Le risposte alle successive domande aperte D5-7 confermano queste tendenze. Infine, le ultime due domande D8-9 poste solo al gruppo VR hanno raccolto valutazioni estremamente positive (il punteggio massimo nella quasi totalità e nessun punteggio inferiore a 4).

## 5. CONCLUSIONI

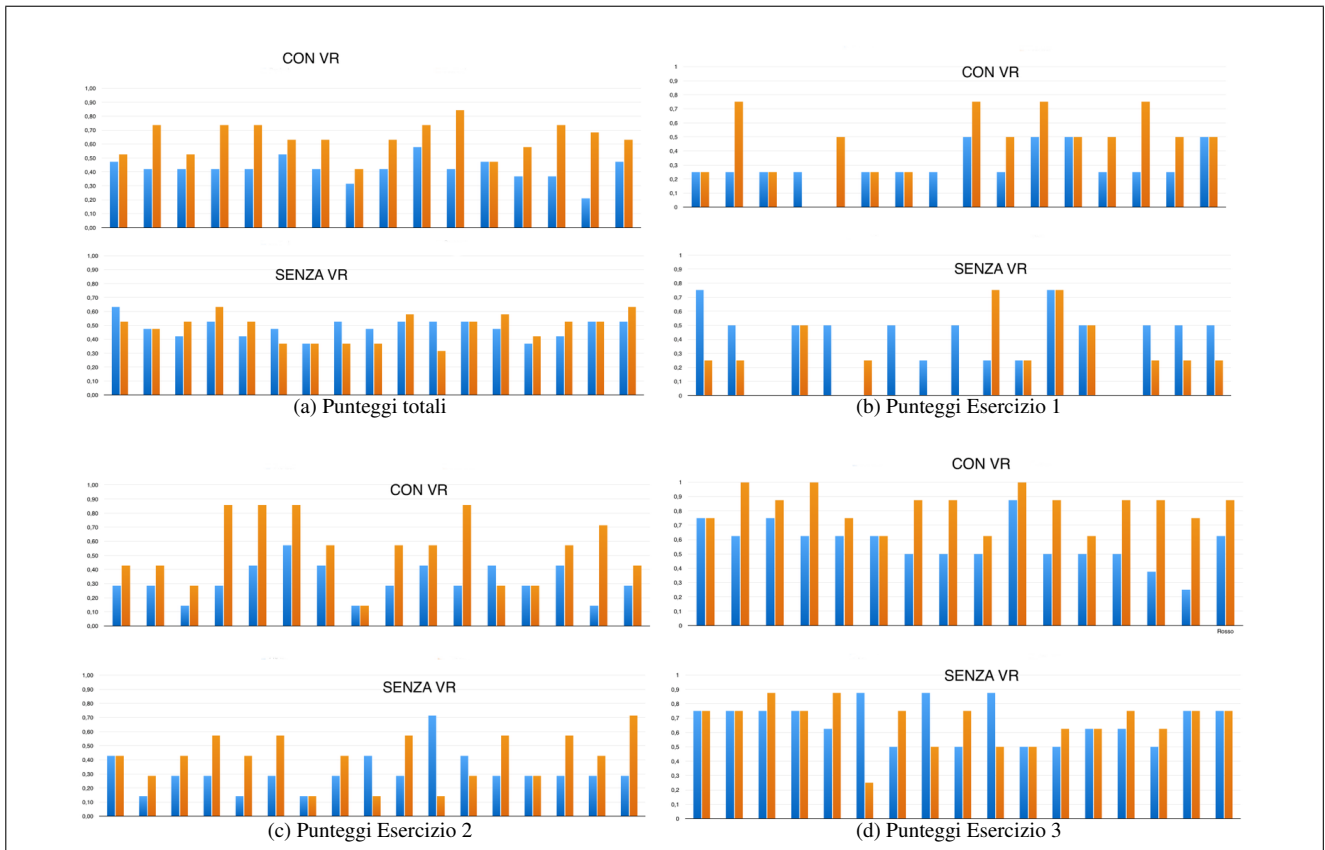
In questo contributo abbiamo presentato il framework multiplatforma VR4EDU, composto da due applicazioni di realtà virtuale in ambiente mobile, MLE e RLE. Abbiamo descritto in dettaglio le scelte di progetto e lo sviluppo delle applicazioni, con particolare attenzione a una innovativa soluzione di "locomozione virtuale" che risulta essere intuitiva e usabile. Abbiamo mostrato che le applicazioni possono funzionare su hardware a basso costo o comunque già posseduto dagli alunni o dalle loro famiglie, diventando quindi appetibili per un uso didattico.

In secondo luogo abbiamo presentato alcuni risultati sperimentali preliminari sull'uso di VR4EDU nella didattica musicale in una scuola primaria. I dati sperimentali sono incoraggianti e suggeriscono che l'uso di VR4EDU come strumento integrativo alla didattica tradizionale porta dei miglioramenti nell'apprendimento degli alunni, e un loro maggiore coinvolgimento. Tuttavia l'analisi di questi risultati è ancora molto incompleta e necessita di approfondimenti volti a verificare se e quanto le differenze osservate tra i due gruppi di alunni (gruppo di controllo e gruppo VR) siano significative dal punto di vista statistico. Sarà quindi necessario esaminare i dati sperimentali attraverso opportuni test statistici, quali ANOVA sui due fattori considerati (metodo didattico e conoscenza pre/post), e successivi test "post-hoc" per confronti appaiati.

Un secondo approfondimento dei dati sperimentali si focalizzerà su un'analisi dedicata agli alunni con BES, per valutare in che misura l'approccio proposto possa portare loro dei vantaggi nel processo di apprendimento, in relazione a specifiche difficoltà personali, di scrittura/lettura o di attenzione.

Sviluppi futuri dell'applicazione riguarderanno in particolare il miglioramento del realismo dell'ambiente virtuale e del livello di dettaglio degli strumenti musicali, nonché una maggiore immersività della scena acustica ad esempio tramite l'uso di registrazioni multitraccia che permettano di distribuire anche acusticamente gli strumenti musicali nello spazio virtuale [35, 36].

Per quanto riguarda i protocolli didattici, sarà possibile estendere il framework VR4EDU al fine di esplorare altri argomenti e materie. Infine, è opportuno notare che negli esperimenti qui riportati l'esperienza di realtà virtuale è stata proposta a pochi alunni alla volta (a causa della disponibilità di soli 3 dispositivi). Di conseguenza la personalizzazione della didattica (spiegazioni fornite a piccoli gruppi per volta, anziché all'intera classe) potrebbe aver influito sul migliore apprendimento. Sarà quindi utile verificare l'efficacia dell'approccio proposto anche nel caso in cui tutti gli alunni usino l'applicazione contemporaneamente.



**Figura 4:** Risultati del pre-test (in blu) e del post-test (in arancione): punteggi globali e punteggi per i singoli esercizi.

## 6. REFERENCES

- [1] M. C. Dwyer, “Reinvesting in arts education: Winning america’s future through creative schools.,” *President’s Committee on the Arts and the Humanities*, 2011.
- [2] R. Črnčec, S. J. Wilson, and M. Prior, “The cognitive and academic benefits of music to children: Facts and fiction,” *Educational Psychology*, vol. 26, no. 4, pp. 579–594, 2006.
- [3] A. Casu, L. D. Spano, F. Sorrentino, and R. Scateni, “Riftart: Bringing masterpieces in the classroom through immersive virtual reality.,” in *Eurographics Italian Chapter Conference*, pp. 77–84, 2015.
- [4] S. Jang, J. M. Vitale, R. W. Jyung, and J. B. Black, “Direct manipulation is better than passive viewing for learning anatomy in a three-dimensional virtual reality environment,” *Computers & Education*, vol. 106, pp. 150–165, 2017.
- [5] T. De Jong, M. C. Linn, and Z. C. Zacharia, “Physical and virtual laboratories in science and engineering education,” *Science*, vol. 340, no. 6130, pp. 305–308, 2013.
- [6] P. Moreno-Ger, J. Torrente, J. Bustamante, C. Fernández-Galaz, B. Fernández-Manjón, and M. D. Comas-Rengifo, “Application of a low-cost web-based simulation to improve students’ practical skills in medical education,” *International Journal of Medical Informatics*, vol. 79, no. 6, pp. 459–467, 2010.
- [7] C. Dede, “Immersive interfaces for engagement and learning,” *science*, vol. 323, no. 5910, pp. 66–69, 2009.
- [8] A. Amer and P. Peralez, “Affordable altered perspectives: Making augmented and virtual reality technology accessible,” in *IEEE Global Humanitarian Technology Conference (GHTC 2014)*, pp. 603–608, Oct. 2014.
- [9] C. Ball and K. Johnsen, “An accessible platform for everyday educational virtual reality,” in *Everyday Virtual Reality (WEVR), 2016 IEEE 2nd Workshop on*, pp. 26–31, IEEE, 2016.
- [10] N. R. Council”, *How people learn: Brain, mind, experience, and school: Expanded edition*. National Academy Press, 2000.
- [11] C. Byrne and T. A. Furness, “Virtual reality and education,” in *Exploring a new partnership: Children, Teachers and Technology*, pp. 181–189, 1994.
- [12] C. Youngblut, “Educational uses of virtual reality technology.,” tech. rep., Institute for Defense Analyses, Alexandria, Va, 1998.
- [13] B. Dalgarno and M. J. W. Lee, “What are the learning affordances of 3-D virtual environments?,” *British*

- Journal of Educational Technology*, vol. 41, pp. 10–32, Jan. 2010.
- [14] Z. Merchant, E. T. Goetz, L. Cifuentes, W. Keeney-Kennicutt, and T. J. Davis, “Effectiveness of virtual reality-based instruction on students’ learning outcomes in k-12 and higher education: A meta-analysis,” *Computers & Education*, vol. 70, pp. 29–40, 2014.
- [15] J. Lave and E. Wenger, *Situated learning: Legitimate peripheral participation*. Cambridge university press, 1991.
- [16] R. C. Schank, T. R. Berman, and K. A. Macpherson, “Learning by doing,” *Instructional-design theories and models: A new paradigm of instructional theory*, vol. 2, pp. 161–181, 1999.
- [17] S. Serafin, C. Erkut, J. Kojs, N. C. Nilsson, and R. Nordahl, “Virtual Reality Musical Instruments: State of the Art, Design Principles, and Future Directions,” *Computer Music Journal*, vol. 40, pp. 22–40, Sept. 2016.
- [18] K. Johanna, G. Romero, D. Andres, R. Lopez, L. A. Luengas, J. Carlos, and B. Guevara, “Virtual flute: Electronic device that uses virtual reality to teach how to play a flute,” in *Education Engineering (EDUCON), 2010 IEEE*, pp. 211–216, IEEE, 2010.
- [19] J. Chow, H. Feng, R. Amor, and B. C. Wünsche, “Music education using augmented reality with a head mounted display,” in *Proceedings of the Fourteenth Australasian User Interface Conference-Volume 139*, pp. 73–79, Australian Computer Society, Inc., 2013.
- [20] X. Xiao, B. Tome, and H. Ishii, “Andante: Walking figures on the piano keyboard to visualize musical motion,” in *NIME*, pp. 629–632, 2014.
- [21] X. Xiao, P. Puentes, E. Ackermann, and H. Ishii, “Andantino: Teaching children piano with projected animated characters,” in *Proceedings of the The 15th International Conference on Interaction Design and Children*, pp. 37–45, ACM, 2016.
- [22] M. Vorländer, *Auralization: Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality*. Springer Publishing Company, Incorporated, 1st ed., 2007. 00379.
- [23] K. Bormann, “Presence and the Utility of Audio Spatialization,” *Presence*, vol. 14, pp. 278–297, June 2005.
- [24] G. Riva, F. Mantovani, C. S. Capideville, A. Preziosa, F. Morganti, D. Villani, A. Gaggioli, C. Botella, and M. Alcañiz, “Affective Interactions Using Virtual Reality: The Link between Presence and Emotions,” *CyberPsychology & Behavior*, vol. 10, pp. 45–56, Feb. 2007.
- [25] A. Gorini and G. Riva, “Virtual reality in anxiety disorders: the past and the future,” *Expert Review of Neurotherapeutics*, vol. 8, pp. 215–233, Feb. 2008.
- [26] H. Sauzéon, P. Arvind Pala, F. Larrue, G. Wallet, M. Déjos, X. Zheng, P. Guittou, and B. N’Kaoua, “The use of virtual reality for episodic memory assessment: effects of active navigation,” *Exp Psychol*, vol. 59, no. 2, pp. 99–108, 2011.
- [27] L. Jäncke, “Music, memory and emotion,” *J Biol*, vol. 7, no. 6, p. 21, 2008.
- [28] M. Geronazzo, A. Bedin, L. Brayda, C. Campus, and F. Avanzini, “Interactive spatial sonification for non-visual exploration of virtual maps,” *International Journal of Human-Computer Studies*, vol. 85, pp. 4–15, Jan. 2016.
- [29] Z. Merchant, E. T. Goetz, W. Keeney-Kennicutt, O.-m. Kwok, L. Cifuentes, and T. J. Davis, “The learner characteristics, features of desktop 3d virtual reality environments, and college chemistry instruction: A structural equation modeling analysis,” *Computers & Education*, vol. 59, pp. 551–568, Sept. 2012.
- [30] F. Steinicke, Y. Visell, J. Campos, and A. Lécuyer, *Human walking in virtual environments*. Springer, 2013.
- [31] M. Usoh, K. Arthur, M. C. Whitton, R. Bastos, A. Steed, M. Slater, and F. P. Brooks Jr, “Walking, walking-in-place, flying, in virtual environments,” in *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, pp. 359–364, ACM Press/Addison-Wesley Publishing Co., 1999.
- [32] W. Hürst and M. Helder, “Mobile 3d graphics and virtual reality interaction,” in *Proceedings of the 8th International Conference on Advances in Computer Entertainment Technology*, ACE ’11, (New York, NY, USA), pp. 28:1–28:8, ACM, 2011.
- [33] K. Bucsuházy and M. Semela, “Case study: Reaction time of children according to age,” *Procedia Engineering*, vol. 187, pp. 408 – 413, 2017. TRANSBALTICA 2017: TRANSPORTATION SCIENCE AND TECHNOLOGY: Proceedings of the 10th International Scientific Conference, May 4–5, 2017, Vilnius Gediminas Technical University, Vilnius, Lithuania.
- [34] M. Wrzesien and M. Alcañiz Raya, “Learning in serious virtual worlds: Evaluation of learning effectiveness and appeal to students in the E-Junior project,” *Computers & Education*, vol. 55, no. 1, pp. 178–187, 2010. 00154.
- [35] G. D. Romigh and B. D. Simpson, “Do you hear where I hear?: isolating the individualized sound localization cues,” *Front. Neurosci*, vol. 8, p. 370, 2014.
- [36] M. Geronazzo, E. Peruch, F. Prandoni, and F. Avanzini, “Improving elevation perception with a tool for image-guided head-related transfer function selection,” in *Proc. of the 20th Int. Conference on Digital Audio Effects (DAFx-17)*, (Edinburgh, UK), pp. 397–404, Sept. 2017.

# TIME IS NOT ON MY SIDE: NETWORK LATENCY, PRESENCE AND PERFORMANCE IN REMOTE MUSIC INTERACTION

**Stefano Delle Monache**  
 Dept. of Architecture and Arts  
 Iuav University of Venice  
 sdellemonache@iuav.it

**Michele Buccoli, Luca Comanducci,  
 Augusto Sarti**  
 DEIB, Politecnico of Milano University  
 name.surname@polimi.it

**Giovanni Cospito, Enrico Pietrocola,  
 Filippo Berbenni**  
 Conservatory of Music G. Verdi of Milano  
 name.surname@consmilano.it

## ABSTRACT

A pilot test on the sense of presence and the quality of the interaction in Networked Music Performance is presented. Subjective measures, based on a presence questionnaire, are combined with objective quality metrics, in order to stress the contribution of temporal factors (i.e., network latency) on the musical experience in the mediated environment. Preliminary results in the scope of chamber music practice are presented.

## 1. INTRODUCTION

Computer-aided musical collaborations between geographically-displaced musicians have been subject of extensive investigation from a variety of perspectives, since the late '90s. Early categorizations of computer systems for musical interaction have been proposed in [1], based on the temporal (synchronous vs. asynchronous) and spatial (co-located vs. remote) dimensions of the performance. The research and development has focused on the technical and perceptual issues (i.e., network delay and audio quality) affecting the on-line, simultaneous performance between musicians located and playing in remote rooms [2].

A substantial body of work has been produced in that area that has been crystallized under the acronym of NMP, namely Networked Music Performance. Gabrielli and colleague provide a valuable picture of the state of the art of NMP research and projects [3, Chapter 2, 3].

From a different angle, a renovated interest in remote collaborative environments has been growing in the area of audio-video (AV) streaming and conferencing systems for educational purposes [4]. NMP technologies and tools are increasingly being available on the market, and proposed as viable standards in blended and distance learning [5].

The EU funded project InterMUSIC<sup>1</sup> (Interactive Environment for Music Learning and Practicing, 2017 - 2020)

<sup>1</sup><http://intermusicproject.eu/>. The Consortium is composed of the Conservatory of Music "G. Verdi" of Milano (Coordinator), the Polytechnic University of Milan, the RDAM Royal Danish Academy of Music of Copenhagen, the LMTA Lithuanian Academy of Music and Theatre of Vilnius, the AEC Association Européenne des Conservatoires, Académies de Musique et Musikhochschulen.

Copyright: ©2018 Stefano Delle Monache et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution License 3.0 Unported](https://creativecommons.org/licenses/by/3.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

aims to bridge the approaches established in NMP research with the opportunities of distance learning and education.

A project-based and practice-led approach to research is aimed at distilling, by the end of the project, an effective and operational environment, and especially a systematized knowledge in the form of best practices and guidelines for the implementation of remote environments for music interaction and education. This involves to author three online pilot courses in music theory and composition, chamber music practice, and vocal training, by means of the implementation of Massive Open Online Courses (MOOCs). With MOOCs, the interaction between teacher and students is usually one-to-many and asynchronous. On the other side, distance learning of the music practice can be delivered through one-to-one synchronous interaction between the teacher and the student.

One of the most popular tool for NMP is designed to provide low-latency interaction by means of dedicated connections and high-performing hardware [6]. In the common scenario, however, students do not have access to such networks and need to connect with the teacher by means of general purpose connections and hardware, which introduce processing and transmission latency.

With regard to NMP, the goal of the project is to investigate the user experience in order to optimize and improve the tools currently available. In this paper, we introduce a study, still in progress, aimed at understanding how temporal factors (i.e., network latency) affect the sense of presence, and the quality of the performance of chamber music duos involved in remote collaboration, i.e music making.

We ask duos to perform a short exercise, under diverse conditions of network delay. The exercise is specifically conceived around musical structures which are functional to pinpointing a set of problems relative to time management, communication mechanisms and mutual understanding between remote performers. A qualitative assessment through questionnaires on the sense of presence and the perceived quality of the performance [7] is combined with quality metrics of the objective performance [8]. A follow-up study will be devoted instead to the investigation of spatial representations, auditory and visual.

The premise is that effective music-making and communication rely on the availability of auditory and visual cues (i.e., sonic gestures) [9], which are inevitably constrained in NMP, and in telepresence environments in general. We turn the traditional engineering approach to NMP research upside down, and seek for design strategies to



**Figure 1:** Instrumentalist positioning in room 1, with frontal view (on screen) of the co-performer displaced in room 2.

compensate, and facilitate a plausible music experience in the mediated environment.

In this respect, we are investigating how simulated network delay, and diverse modes of audio-visual spatial representation, separately, affect the subjective experience of being present and together in the shared reality environment. Sensory breadth and depth, degree of control and anticipation of events, together with the overall interactivity of the environment, represent crucial elements in both presence and performance, being the first a prerequisite for the second [10].

The paper is organized as follows: in Section 2 we provide the background context on top of which we are building our study; in Section 3 we elaborate on the concept of presence in the embodied cognition framework; Section 4 introduces the pilot experiment, and the methodology; we discuss the preliminary observations collected, in Section 5.

## 2. BACKGROUND

Both remote music performance and tuition face two major issues, that is dealing with temporal and spatial factors that are crucial for the effective and expressive performers and tutor-student communication.

Whereas remote music performance demands instantaneous and immediate responses, musical interaction in remote tuition rather reflects the flexibility of turn-taking in conversations [11]. In both situations, visuo-spatial forms of communication and behaviors are exploited to support mutual understanding and negotiation between peers.

Properly designed spatial elements in remote interactive environments may facilitate the compensation of time-dependent misalignments in the communication. In this respect, the design of the stage and the performance scene implies not only the physical displacement of various equipments in the rooms, but also the choice of appropriate solutions in both visual and sound feedback display [4, 11]. Various off-the-shelf strategies, such as the use of spatialized audio to increase the sense of presence [12], life-size and near-life-size visual display to preserve a coherent distance perception in virtual proximity [13], projections and video loop techniques to support synchronous interaction [14], have been proposed. These aspects are not considered in the current experimental research and will be subject of separate investigation.

In the following subsections we illustrate some relevant

network latency effects on remote music performance, and discuss the available environments and tools, from the InterMUSIC perspective.

### 2.1 Temporal factors in remote interaction

Temporal factors refer to latency issues in data transfer, inherent to any NMP due to many factors, which can be broadly divided into processing time and network latency [3, Chapter 3]. The former includes the time for digital-to-analog and analog-to-digital conversion, for the acquisition and rendering of AV signals, their buffering and their packetization/depacketization for transmission. The latter is the latency introduced by the transmission of the packets, and it depends on the properties of the network connection and of the streams to be transmitted [2].

The temporal factors in remote interaction affect the quality of the music performance, and require musicians to cope with the delayed interaction [15]. Music tolerance to network latencies have been assessed in a variety of conditions, and quantified in a wide range of delay thresholds [2]. In [8], Rottondi and colleagues conducted a set of experiments on NMP and performed a signal-based analysis of the recorded performances to draw some conclusions on the effect of the latency and of other musical factors on the quality of the performance. For instance, they showed that adverse network conditions (20 – 60 ms) affect the tempo of the performance, by inducing a pronounced tendency to deceleration, computed as explained in Section 4.4.

Carôt and Werner identify different strategies that musicians can follow to cope with the latency, depending on the style and the tempo of the music performance at hand [16]. The best scenario, mostly feasible for low latencies, is the Realistic Interactive Approach, when the performers flawlessly interact together. With latencies higher than 25ms, musicians tend to follow a Master-Slave or a Laid-Back Approach. In the former, one musician leads the performance and others follow her lead; in the latter, one musician plays slightly behind the rhythm to compensate for the delay. These approaches are also common in co-presence performances, when a performer plays a lead role or behind the rhythm.

Jointly with latency, other factors have been observed to affect the quality of a NMP. Barbosa and Cordeiro observe that slow instruments' attack times lead to higher deceleration, but also higher synchronization between musicians [17], and Rottondi and colleagues observe a correla-

tion between deceleration and pieces' rhythmic complexity and instruments' noisiness [8]. Rhythmic and timbral factors will be investigated in future experiments.

## 2.2 Environments for remote interaction

In the scope of the InterMUSIC project, we intend to use LOLA<sup>2</sup>, developed by Conservatorio di Musica Giuseppe Tartini of Trieste, in collaboration with GARR Consortium [6], and UltraGrid<sup>3</sup>, developed by CESNET's Laboratory of Advanced Networking Technologies (Sitola) [18].

LOLA requires to use dedicated hardware and the European National Research and Education Network to provide low-latency, large bandwidth connection. LOLA runs as a Windows program with a GUI, and offers several codecs for AV signals, AV recording functionality, and a set of tools for remote assistance. Due to its features, LOLA has been widely used for artistic remote performances [6]. However, LOLA is designed to work in best conditions, and it is not open-source for the community.

UltraGrid is a cross-platform program which also implements several AV codecs. Since it is based on command line, it is more difficult to use for non-tech-savy users. The code of UltraGrid is public<sup>4</sup>, hence it is possible to implement additional functionalities, e.g., the support for spatial factors. UltraGrid can reach an end-to-end latency as low as 50 ms, which may be too high for some scenarios.

In this pilot test, we use LOLA because of its low-latency. We intend to use UltraGrid in future experiments for spatial or other factors.

## 3. PRESENCE: KEEPING THE GOOD COMPANY

Network constraints and high quality sound reproduction have been considered the main factors affecting the plausibility of music interaction in networked performance. Research strictly focused then on the sensory and control dimensions involved in the experience of playing together remotely [19].

In InterMUSIC, we approach the problem of providing an effective communication and interaction environment for networked music practice and teaching, by looking at presence studies [20, 10] in the embodied cognition framework [21, 7]. The basic assumption is that the experience of music emerges in interaction, as complex network of predictive models of observable patterns and intentional states, that are acquired through knowledge and skills [22].

The concept of presence is subject of ongoing investigation as potential construct from which deriving appropriate measures of the effectiveness of virtual environments. Several theoretical and operational models have been proposed over the last two decades (see [20] for a comprehensive survey), also in the attempt to linking such a construct to task performance. A conclusive demonstration has not been provided yet [10].

In the scope of InterMUSIC research, we adopt the definitions of presence proposed in [20, 7], that is presence

<sup>2</sup> LOLA (LOW LATency audio video streaming system) <https://lola.conts.it>

<sup>3</sup> UltraGrid <https://lola.conts.it/>

<sup>4</sup> UltraGrid code <https://github.com/CESNET/UltraGrid>

as *the cognitive feeling of being in a place*, intended as *the perceived realness of the mediated experience*. In this study, we refer to the presence model by Schubert and colleague [7], emerging from the factor analysis of 246 answers to a 75-item survey of questions taken from several questionnaires. Due to space constraints, we do not elaborate further on this definition, instead we discuss the basic components of the presence experience, as operational constructs that have been assessed experimentally in established questionnaires [20].

The sense of presence is tied to action in the (mediated) environment, and results from the interpretation of the mental model derived. Given the focus of the chamber music practice task, the suppression of conflicting sensory stimuli and the focused allocation of attentional resources is of primary importance.

## 3.1 Components of the presence experience

Presence encompasses three major categories of constructs: i) the subjective experience of being there (spatial presence, involvement, realness); ii) evaluations of the immersive technology (interface awareness and quality); iii) evaluations of the interaction (predictability/anticipation) [7].

**Spatial presence** refers to the emerging relation between the mediated environment as a space and the instrumentalist's own body. **Involvement** or flow retains the attention side of the presence experience, that is the relative concentration and focus on the real and the mediated environments, in terms of keeping the compatible information from the real environment and suppressing the incompatible one. **Realness** encompasses reality judgements with respect to the meaningfulness and coherence of the experience with the expectations from the real-world.

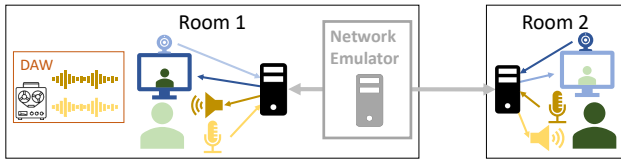
**Interface awareness and quality** take in account distraction factors, and in general provides a clue of the mastery of the interface in the specific activity at hand. As additional construct, **quality of immersion** considers the vividness, coherence and fidelity of sensory factors (i.e., auditory and visual cues, and valid actions) as supported by the technology.

**Predictability** refers to the possibility to anticipate what will happen next, in terms of activation of motor representations as a consequence of perceiving while playing.

The peculiarity of NMP, and in particular in the context of chamber music practice in InterMUSIC, makes the remote interaction environment very constrained, since the music making task is very specific and demanding, in terms of active survey. Put in design terms, this represents an advantage, potentially providing clearer user requirements. Much of expected outcomes are aimed at collecting and organizing the instrumentalists expectations.

## 4. THE EXPERIMENTAL STUDY

The pilot experiment took place at the Conservatory of Music "G. Verdi" of Milano, in two dedicated rooms, equipped with direct network connection and all the necessary infrastructures. How does the experience of network



**Figure 2:** Setup of the pilot experiment, where two musicians perform together through a Network Emulator.

| Room | Couple A | Couple B  | Couple C   | Couple D | Couple E |
|------|----------|-----------|------------|----------|----------|
| 1    | mandolin | accordion | percussion | harp     | alto sax |
| 2    | mandolin | guitar    | percussion | flute    | alto sax |

**Table 1:** Instrumentation and location of the instrumentalists in room 1 or room 2.

delay and interruption affect the involvement and realness, the predictability, the interface awareness and quality, and the overall quality of the musical performance? The same question applies separately to AV spatial representations in remote interaction and will be investigated in a forthcoming study. The outcomes are expected to provide relevant design implications in the staging of classrooms dedicated to remote music practice and teaching.

In Section 4.1 we describe the setup of the experiment, and in Section 4.2 the participants and the stimuli for the experiment. Performances' quality and sense of presence are assessed through subjective questionnaire, reported in Section 4.3, and objective metrics, explained in Section 4.4. The scores of the stimuli and the presence questionnaire are publicly available<sup>5</sup>. A further research goal in the long term is to provide evidence of links between presence and performance in remote music interaction.

### 4.1 Apparatus

Figure 2 represents the experimental setup. Two musicians perform in two rooms with no direct sound interferences.

Musicians interact by means of contact microphones (monaural acquisition), loudspeakers (monaural rendering), cameras and screens. The spatial arrangement of visual displays and loudspeakers reflects the current practices established in remote music practice and tuition, that is a frontal view of the co-performer in order to improve eye contact [6, 23]. Figure 1 shows the staging in Room 1.

The hardware equipment is connected to two Windows computers, which are two high-end Intel/Nvidia powered workstations with i7 esa/octa core processors, using PCIe audio cards, according to both LOLA and UltraGrid hardware and software requirements. The computers communicate together through a Gigabit ethernet connection by means of a server with two Ethernet interfaces acting as a Network Emulator to add a fixed delay to both audio and video streams<sup>6</sup>. The effect of jitter will be considered in future experiments [8]. The server is placed in Room 1 to be easily accessible during the tests and to ease troubleshooting in case of network issues.

Audio output and input of the performer in Room 1 is redirected to a Digital Audio Workstation to record the per-

<sup>5</sup> <https://tinyurl.com/intermusicCIM2018>

<sup>6</sup> Using Linux command tc

| Ex. | Rhythm       | Melody             | Expression (dynamics, articulation, agogic) |
|-----|--------------|--------------------|---|
| 3a  | homorhythm   | octave or unison   | static                                      |
| 8a  | homorhythm   | opposite direction | static                                      |
| 5   | heterorhythm | opposite direction | alternation                                 |
| 7   | heterorhythm | homodirection      | climax                                      |
| 6b  | phasing      | octave or unison   | static                                      |
| 4b  | slicing      | slicing            | dynamic                                     |
| 2b  | imitation    | imitation          | imitation                                   |
| 3b  | ostinato     | pedal tones        | static                                      |

**Table 2:** Types of musical structures in rhythm-melody-expression relationship.



**Figure 3:** Example of homorhythmic, unison melody, with static expression. Left: Flute and harp. Right: Wood blocks and tom-toms.

| Condition          | 1  | 2  | 3  | 4  | 5  | 6   |
|--------------------|----|----|----|----|----|-----|
| Delay (2-ways, ms) | 28 | 33 | 50 | 67 | 80 | 134 |

**Table 3:** Latency values in the six conditions.

formance (from the perspective of Room 1). These recordings are used to compute content-based metrics of quality, as described in Section 4.4.

### 4.2 Participants and task

Ten volunteers (five duos, five males, five females, age ranging from 14 to 29, average age = 21,9 years, SD = 4,7) were recruited from the class of chamber music practice. They are all musicians with at least five years of academic musical practice. Each duo had already a certain familiarity of minimum two weeks of rehearsal. Table 1 shows the instrumentation and respective location per each duo in rooms 1 and 2 (see Figure 2).

The stimuli were designed in order to encompass diverse basic structures of musical interaction with respect to time management and communication strategies. The rationale of the stimuli (i.e., the scores proposed to the duos) concerns simple, yet constraining aspects of synchronicity in musical time, as established in western music tradition, that is the tight link between the musical dimensions of rhythm, melody, and expression.

In this respect, we looked at Béla Bartók's *Mikrokosmos* piano pieces [24], which represent a valuable methodological compendium of exercises in meaningful *rhythm-melody-expression* relationships. In Table 2, we pinpoint eight types of structures per musical dimension. These can be combined in diverse expressive relationships of musical synchronicity. A few examples are reported and referenced in the fourth column to guide the reader in the understanding of the score.

Figure 3 shows an example of homorhythmic, unison melody, with static expression relationship respectively ex-

tracted from the scores for flute and harp (left) and percussions (right). The musical stimuli have a duration of 3 minutes, and a reference tempo of 112 BPM.

Each duo had to perform the exercise, under six different conditions of emulated network delay, reported in Table 3. As described in Section 2.1, a certain latency is unavoidable due to time for audio processing. By setting the network latency to 0 ms, we estimated the two-way processing time as 28 ms, which is the first condition in Table 3. We use two sets of latency values: a deterministic set, and a set corresponding to rhythmic figures, given the reference tempo. Specifically, 33, 67 and 134 ms correspond to one 64th, 32th or 16th respectively. The idea behind this choice is to investigate whether having a musically-meaningful latency will help performers to cope with it. However, due to the few examples, statistical tests in this regard have not been conducted yet.

The sequence of six conditions was randomized for each duo. Before each session, each duo was briefed in room 1, and the task was introduced, within the scope of InterMUSIC, as that of performing the exercise for 6 times, in physically displaced locations, and through the telepresence environment. No information about the nature of the six network delay conditions was disclosed. The score of the exercise was explained and handed out. Participants were informed about the duration of the exercise and the approximately overall duration of the experimental session (90 min.), and introduced to the questionnaire on presence. They were asked to fill in the 5-item questionnaire after each single repetition, and the general 27-item questionnaire at the end of the session. These are extensively reported in Section 4.3. Further comments were collected at the end of the test.

After the brief, the musicians settled in their respective room, and a 15-minute rehearsal was devoted to adjust their positioning, framing and volume levels in order to provide a comfortable environment. In addition, they could rehearse and get acquainted with the score.

### 4.3 Presence questionnaire

The questionnaire used in the study was constructed by merging three reference questionnaires on presence [25, 7, 26], and selecting the most appropriate items with respect to the music performance task. Items were partially rephrased and adapted to the musical context and language. The resulting close-ended, 7-point likert scale questionnaire was edited in Italian.

The questionnaire is split in two main parts: a general post-experiment 27-item questionnaire subdivided into five groups; and a post-repetition questionnaire with five questions extracted from the general one. We report the post-experiment questionnaire in Table 4 (with the median, mean and standard deviation of the answers).

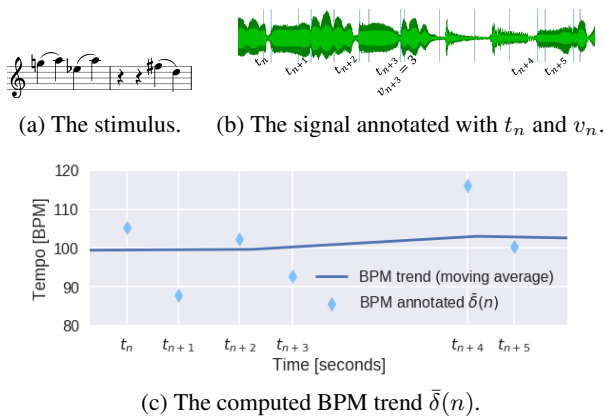
The questionnaire is constructed around three main concepts: i) Presence (G1); ii) Immersion (G2, G3, G4); iii) Quality of the Performance (G5). G1 on presence is further composed of items on three main constructs, that is *spatial presence* ( Q1.2-4), *involvement* (Q1.3, Q1.6-9), *realness* (Q1.1, Q1.5). G2 encompasses items on *predictability and*

| Question   | Median | Mean | Std  |
|--|--------|------|------|
| <b>1) Involvement and realness</b>   |        |      |      |
| 1.1) How much did your musical experience in the remote environment seem consistent with your real world experiences?    | 4.0    | 4.40 | 1.36 |
| 1.2) In the remote environment I had a sense of "being there".   | 3.5    | 4.17 | 1.77 |
| <i>1.3) The sense of playing in the remote environment was compelling.</i>   | 5.0    | 4.50 | 1.50 |
| 1.4) I had a sense of playing in the remote environment, rather than performing something from outside.                  | 4.0    | 3.80 | 1.17 |
| 1.5) How realistic did the remote environment seem to you?   | 4.5    | 4.33 | 0.75 |
| 1.6) How aware were you of the real world surroundings around you during the performance?                                | 4.5    | 4.00 | 1.91 |
| 1.7) How completely were you able to actively survey the musical environment using vision?                               | 4.0    | 3.83 | 1.57 |
| 1.8) How completely were you able to actively survey the musical environment using audition?                             | 5.0    | 4.67 | 0.94 |
| <i>1.9) The delay affected the sense of involvement.</i>   | 4.5    | 4.00 | 1.63 |
| <b>2) Predictability and Interaction</b>   |        |      |      |
| 2.1) The musical interaction in the remote environment seemed natural.   | 5.0    | 5.00 | 0.82 |
| 2.2) The environment was responsive to actions that I performed.   | 4.0    | 4.00 | 1.83 |
| <i>2.3) I was able to anticipate the musical outcome in response to my performance in the remote environment.</i>        | 5.0    | 5.00 | 1.00 |
| 2.4) The environment was responsive to actions performed by my partner.  | 4.5    | 4.25 | 1.48 |
| 2.5) I was able to anticipate the musical outcome in response to the performance by my partner in the remote environment | 4.0    | 3.83 | 1.21 |
| 2.6) It was difficult to cope with the distance performance.   | 4.0    | 3.83 | 1.57 |
| <b>3) Interface awareness and quality</b>  |        |      |      |
| 3.1) How well could you concentrate on the music performance rather than on the mechanisms required to perform?          | 5.5    | 5.00 | 1.53 |
| 3.2) How aware were you of the display and control devices/mechanism?  | 4.0    | 4.00 | 1.10 |
| 3.3) How much did the visual display quality interfere or distract from performing?                                      | 3.0    | 3.17 | 1.21 |
| 3.4) How much did the auditory display quality interfere or distract from performing?                                    | 6.0    | 5.17 | 1.46 |
| <i>3.5) How much delay did you experience between your actions and expected outcomes?</i>                                | 4.5    | 4.33 | 1.60 |
| <b>4) Quality of the immersion</b>   |        |      |      |
| 4.1) The visual representation made me feel involved in the remote environment.  | 3.0    | 3.33 | 0.94 |
| 4.2) The auditory representation made me feel involved in the remote environment.  | 4.0    | 4.00 | 1.73 |
| 4.3) I felt involved in the remote environment experience.   | 5.0    | 4.60 | 0.49 |
| <b>5) Quality of the music performance</b>   |        |      |      |
| 5.1) How quickly did you adjust to the experience of playing in the remote environment?                                  | 4.5    | 4.17 | 1.67 |
| 5.2) It was easy to cope with the delay to adjust the quality of the performance.  | 3.5    | 3.33 | 1.25 |
| 5.3) How proficient in remote music playing did you feel at the end of the experience?                                   | 5.0    | 4.67 | 1.89 |
| <i>5.4) The delay affected the quality of my performance.</i>  | 5.0    | 4.33 | 1.80 |

**Table 4:** Statistics of the answers to the post-experiment questionnaire defined in Sec. 4.3. We indicate in italics the post-repetition questions.

*interaction* (Q2.x), G3 on *interface awareness and quality* (Q3.x). G4 focuses on the *quality of immersion* (Q4.x), and G5 is dedicated to the subjective assessment of the *quality of the performance* (Q5.x).

The post-repetition questionnaire contains five items (indicated in italics in Table 4) referring to the three concepts, with a special emphasis on the experience of delay.



**Figure 4:** Representation of the annotation procedure for computing objective content-based metrics.

#### 4.4 Objective quality metrics

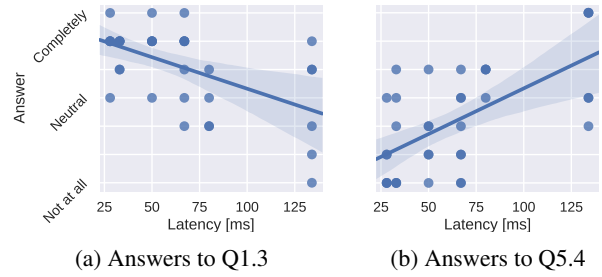
Beside the subjective questionnaire, we conducted a set of content-based analysis on the recordings of the study. The goal of this analysis is to obtain an objective metrics of the quality of the performance. In the literature, researchers have proposed different metrics, related to the rhythmic trend of the performances [2]. Fig. 4 provides an example of the annotation procedure that leads to the computation of these metrics. Given the stimulus represented by a score (Fig. 4a), the first step is to annotate, in the recordings, the instants when an onset occurs on the beat, as  $t_1, t_2, \dots, t_N$  (Fig. 4b). In order to address the issues of beats occurring without an onset, e.g., because of a four-quarter note, we also annotated the amount of beats occurring between the  $n$ -th instant and the following ( $t_n$  and  $t_{n+1}$ ) as  $v_n$ . In Fig. 4b, for example, the onset occurring at  $t_{n+3}$  is followed by a two-quarter pause, hence the onset at  $t_{n+4}$  occurs after three beats and  $v_{n+3} = 3$ . We convert the set of annotations ( $t_1, v_1$ ), ..., ( $t_N, v_N$ ) to the tempo samples (in BPM)  $\bar{\delta}(n) = (60 \cdot v_n) / (t_{n+1} - t_n)$ .

Following the previous example, we show the final BPM annotation in Figure 4c. We apply a 5-second moving average filter to  $\bar{\delta}(n)$  to compensate the variance due to the musical agogics and the resulting imprecision of the manual annotation, as shown in Fig. 4c.

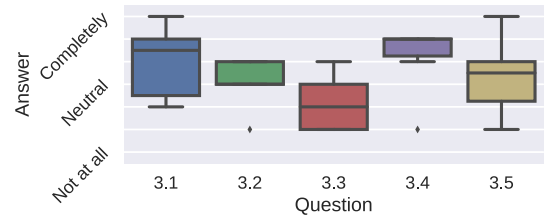
From on  $\bar{\delta}(n)$  we can compute several metrics [2] related to the symmetry between two performers, or the imprecision/regularity of the execution, or the slope of the linear approximation of  $\bar{\delta}(n)$ . The latter provides a compact descriptor of the trend of the tempo, since  $\kappa = 0$  when the tempo remains steady for the whole performance and it assumes positive or negative values in case of acceleration or deceleration, respectively. While in the InterMUSIC project we intend to use all the aforementioned metrics, in the rest of this paper we will focus on the latter, due to its interpretability and graphical representation.

## 5. RESULTS AND DISCUSSION

Given the limited number of collected sessions, we provide the reader with a narrative of the information that we are able to extract from both subjective and objective evalua-



**Figure 5:** Answers to Q1.3 and Q5.4 in the post-repetition questionnaire with respect to the latency condition.



**Figure 6:** Distribution of the answers to the questions related to G3 *Interface awareness and quality* (see Table 4).

tion. We consider only three sessions out of five, since two sessions were either not completed or deeply biased.

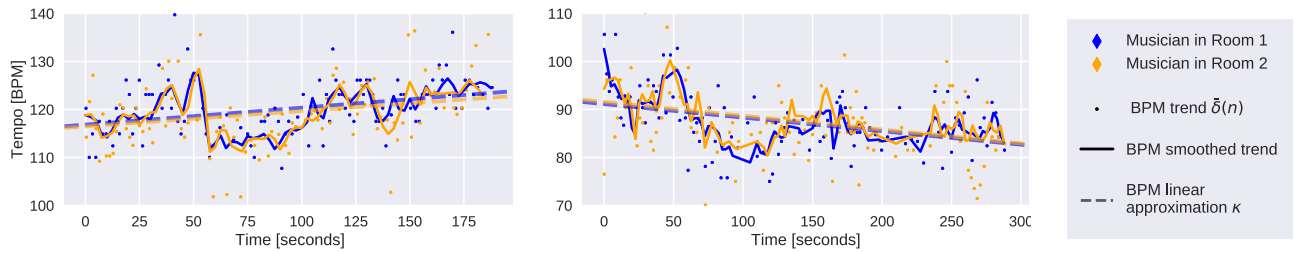
#### 5.1 Subjective evaluation

The subjective evaluation is aimed at measuring the general sense of presence in NMP performance, and at observing whether and how delay affects the presence components, that is the attention allocation to the real and mediated environment, the coherence and fidelity of the overall experience, in addition to the perceived quality of the performance.

Answers to the post-repetition questionnaire return predictable, yet interesting results. For instance, in Fig. 5 we show the answers to two post-repetition questions on involvement and the subjective experience of delay: Q1.3 *The sense of playing in the remote environment was compelling* (Fig. 5a) and Q5.4 *The delay affected the sense of involvement* (Fig. 5b). The answers' distributions highlight a negative effect of latency levels to musicians' involvement in the environment.

From the answers to the post-repetition questionnaire in Table 4, it emerges a general distress caused by the latency and a willingness to adapt and find ways to cope with it. In post-experiment comments, participants C1 and D2 reported a lack of “musical connectedness”, despite the plausibility of the experience, which was also accounted by participants C2 and E1. Of interest, the decrease in involvement or flow, due to longer delays, resulted in the difficulty or impossibility to understand who or which was responsible of playing out of time.

Focusing on G3 *interface awareness and quality*, we show the distribution of answers in Fig. 6. The median values of the answers suggest that musicians feel confident to concentrate on the musical task, adapt and master the interface at hand. It is worth noting that the visual display quality seems not to interfere or distract from the perform-



(a) Signal-based metrics of the second repetition of couple C, with 134 ms latency.

(b) Signal-based metrics of the fifth repetition of couple D, with 134 ms latency.

(c) Legend of the figures

**Figure 7:** BPM trend and its linear approximation for two sets of repetitions.

ing (Q3.3), while the audio quality does (Q3.4). We believe this is due to the poor quality of immersion of the visual representation (Q4.1), which makes it difficult for the performers to rely on vision to actively survey the performance (Q1.7). It is possible that this condition led the musicians to rather rely on the audio feedback in their performance. As a general comment, all the couples reported unanimously that the frontal screen resulted in a less natural interaction, since they normally use peripheral vision to oversee the co-performers on their side.

## 5.2 Objective evaluation

We compute the tempo trend  $\bar{\delta}(n)$  for all the recordings of the experiments. In Figure 7 we show two sets of annotations and corresponding tempo trends (scatter plot), smoothed trend (continuous lines) and linear approximation computed from  $\kappa$  (dashed lines) for musicians in room 1 (blue) and 2 (orange).

Figure 7a shows the results regarding the two percussionists (couple C), playing with a latency of 134 ms. The smoothed trend shows that there is a high degree of synchronization between the performers, and the trends are highly correlated, making them to increase the tempo during the performance. This is probably due to the fact that with such a latency the musicians were not able to follow each other and instead opted for a master-slave approach. The percussionists performance was particularly challenging because of the nature of their instruments that caused a high audio feedback, which further distracted them from the performance.

Figure 7b shows the results for the harp and flute players (couple D) playing with a latency of 134 ms. While the latency value is the same as in Figure 7a, the results are dramatically different. By listening to the recordings, we note that the two instrumentalists attempt to perform with a realistic interaction approach, but the prohibitive latency condition led them to strongly slow down over time with respect to the reference tempo.

Our analysis of the compensation strategies followed by the two couples are further supported by the comments that they provided after the end of the session. While the percussionist C2 stated that she focused on the internal tempo, ignoring the partner's delayed performance, both D1 and D2 stated that they were trying to follow each other's performance.

The discrepancies in the two performances are also due

to the inherent difference in the musicians' instruments. Percussionists are able to efficiently follow themselves even at higher levels of latency [13], which is shown by the relative similar smooth fits of the tempo trend of both musicians, as shown in Fig. 7a. D1, the harpist, and D2, the flautist, must confront bigger challenges due to the constrictive relationship imposed both by the melodic and agogic constraints of their stimuli. Unlike the percussionist, who have to focus mainly on the tactus, the harpist and the flautist have to preserve a synchronicity in pitch as well.

## 6. CONCLUSIONS

NMP practices are subject of renovated interested in the wider scope of music teaching. Within the activities of the InterMUSIC project, we are currently tackling the most demanding aspects of music interaction in remote environments. We seek to gain a wider understanding of NMP issues in the context of chamber music, and provide design solutions that cope with temporal and spatial factors affecting the remote musical experience.

In this work, we presented the design of a pilot experiment, which included the technical setup, the creation of stimuli and presence questionnaire and the introduction of objective metrics of quality for performance. While in the first experiment we focused on temporal factors, in future work we plan to collect a valuable number of sessions and start exploring diverse design strategies in spatial representations. This will allow us to address issues raised by the participants of the experiments, such as the use of peripheral vision or the need of spatialized audio. In addition, we intend to take a wider variety of objective metrics into consideration [2], as well as to introduce physiological measures of presence [20].

## 7. ACKNOWLEDGMENT

This study was conducted for the InterMUSIC project, which received the financial support of the Erasmus+ National Agency under the KA203 Strategic Partnership action under grant number: 2017-1-IT02-KA203-036770.

The authors would like to thank Daniele Moro, Sebastian Troia and Marco Fumagalli for their help in the setup of the Network Emulator. We would also like to thank the participants of the pilot test for their effort and their availability to be contacted for further experiments.

## 8. REFERENCES

- [1] Á. Barbosa, “Displaced soundscapes: A survey of network systems for music and sonic art creation,” *Leonardo Music Journal*, pp. 53–59, 2003.
- [2] C. Rottondi, C. Chafe, C. Allocchio, and A. Sarti, “An overview on networked music performance technologies,” *IEEE Access*, vol. 4, pp. 8823–8843, 2016.
- [3] L. Gabrielli and S. Squartini, *Wireless Networked Music Performance*. Springer, 2016.
- [4] F. Alpiste Penalba, T. Rojas-Rajs, P. Lorente, F. Iglesias, J. Fernández, and J. Monguet, “A telepresence learning environment for opera singing: distance lessons implementations over internet2,” *Interactive Learning Environments*, vol. 21, no. 5, pp. 438–455, 2013.
- [5] M. Iorwerth, D. Moore, and D. Knox, “Challenges of using networked music performance in education,” in *AES Conference:UK 26th Conference: Audio Education (August 2015)*, 8 2015.
- [6] C. Drioli, C. Allocchio, and N. Buso, “Networked performances and natural interaction via lola: Low latency high quality a/v streaming system,” in *Information Technologies for Performing Arts, Media Access, and Entertainment*, pp. 240–250, Springer, 2013.
- [7] T. Schubert, F. Friedmann, and H. Regenbrecht, “The experience of presence: Factor analytic insights,” *Presence: Teleoperators & Virtual Environments*, vol. 10, no. 3, pp. 266–281, 2001.
- [8] C. Rottondi, M. Buccoli, M. Zanoni, D. Garao, G. Verticale, and A. Sarti, “Feature-based analysis of the effects of packet delay on networked musical interactions,” *J. Audio Eng. Soc.*, vol. 63, no. 11, pp. 864–875, 2015.
- [9] R. I. Godøy and M. Leman, *Musical gestures: Sound, movement, and meaning*. Routledge, 2010.
- [10] E. B. Nash, G. W. Edwards, J. A. Thompson, and W. Barfield, “A review of presence and performance in virtual environments,” *International Journal of human-computer Interaction*, vol. 12, no. 1, pp. 1–41, 2000.
- [11] S. Duffy, P. G. Healey, *et al.*, “Co-ordinating non-mutual realities: The asymmetric impact of delay on video-mediated music lessons,” in *Proc. of the 39th Annual Conference of the Cognitive Science Society*, 2017.
- [12] C. Hendrix and W. Barfield, “The sense of presence within auditory virtual environments,” *Presence: Teleoperators & Virtual Environments*, vol. 5, no. 3, pp. 290–301, 1996.
- [13] J. R. Cooperstock, “Interacting in shared reality,” in *Proc. HCI International, 11th Int. Conf. on Human-Computer Interaction*, 2005.
- [14] C. Knudsen, “Synchronous virtual spaces—transparent technology for producing a sense of presence at a distance,” in *Proc. of The 2001 Telecommunications for Education and Training conference, TET*, 2001.
- [15] C. Chafe and M. Gurevich, “Network time delay and ensemble accuracy: Effects of latency, asymmetry,” in *Audio Engineering Society Convention 117*, Oct 2004.
- [16] E. Carôt and C. Werner, “Network music performance - problems, approaches and perspectives,” in *in International School of new Media, Institute of Telematics, University of Lbeck. Music in the Global Village - Conference*, 2007.
- [17] A. Barbosa and J. Cordeiro, “The influence of perceptual attack times in networked music performance,” in *Audio Engineering Society Conference: 44th International Conference: Audio Networking*, Nov 2011.
- [18] P. Holub, L. Matyska, M. Liška, L. Hejtmánek, J. Denemark, T. Rebok, A. Hutanu, R. Paruchuri, J. Radil, and E. Hladká, “High-definition multimedia for multi-party low-latency interactive communication,” *Future Generation Computer Systems*, vol. 22, no. 8, pp. 856–861, 2006.
- [19] W. Woszczyk, J. Cooperstock, J. Roston, and W. Martens, “Shake, rattle, and roll: Getting immersed in multisensory, interactive music via broadband networks,” *Journal of the Audio Engineering Society*, vol. 53, no. 4, pp. 336–344, 2005.
- [20] R. Skarbez, F. P. Brooks Jr, and M. C. Whitton, “A survey of presence and related concepts,” *ACM Computing Surveys (CSUR)*, vol. 50, no. 6, p. 96, 2017.
- [21] T. W. Schubert, “A new conception of spatial presence: Once again, with feeling,” *Communication Theory*, vol. 19, no. 2, pp. 161–187, 2009.
- [22] M. Leman, P.-J. Maes, L. Nijs, and E. Van Dyck, “What is embodied music cognition?,” in *Springer Handbook of Systematic Musicology* (R. Bader, ed.), pp. 747–760, Springer, 2018.
- [23] S. Duffy and P. Healey, “A new medium for remote music tuition,” *Journal of Music, Technology & Education*, vol. 10, no. 1, pp. 5–29, 2017.
- [24] B. Bartók, *Mikrokosmos, 153 Progressive Piano Pieces, New Definitive Edition*. London: Boosey & Hawkes Music Publishers Limited, 1987.
- [25] B. G. Witmer and M. J. Singer, “Measuring presence in virtual environments: A presence questionnaire,” *Presence*, vol. 7, no. 3, pp. 225–240, 1998.
- [26] G. Robillard, S. Bouchard, P. Renaud, and L. Cournoyer, “Validation canadienne-française de deux mesures importantes en réalité virtuelle: l’immersive tendencies questionnaire et le presence questionnaire,” *25e Congrès Annuel de la Société Québécoise pour la Recherche en Psychologie (SQRP)*, Trois-Rivières, 2002.

# Deep Learning per l'Elaborazione Digitale del Suono: possibilità di integrazione nella prassi compositiva elettroacustica

**Martino Sarolli**

Conservatorio G. Verdi, Torino  
mmsarolli@gmail.com

**Andrea Agostini**

Conservatorio G. Verdi, Torino  
andrea.agostini@conservatoriotorino.eu

## ABSTRACT

Il termine *deep learning* fa generalmente riferimento ad un repertorio di tecniche di *machine learning* basate sulle reti neurali artificiali multilivello (*deep neural networks*), protagoniste del recente proliferare di nuove applicazioni dell'intelligenza artificiale in ambiti sempre più vasti e diversificati.

Aspetto chiave e terreno comune, l'aggettivo *deep* indica la presenza, all'interno delle reti neurali, di *layer* multipli che elaborano più livelli di astrazione, in modo da poter esprimere rappresentazioni complesse in termini di rappresentazioni più semplici.

Le loro capacità di controllo su livelli di complessità superiori rendono i sistemi *deep learning* in grado rappresentare e gestire dati ad alto livello di astrazione come, in ambito sonoro, si possono considerare le proprietà timbrico-spetttrali e dinamiche di un suono.

Potendo accedere ed operare sulle rappresentazioni delle caratteristiche sonore create all'interno delle *deep neural networks*, è possibile sviluppare nuovi processi di analisi, elaborazione e sintesi.

Tra le prime implementazioni disponibili di queste nuove possibilità verranno descritti struttura e funzionamento del sistema Nsynth e proposto un tentativo di integrare alcune delle tecniche che mette a disposizione all'interno del processo di realizzazione di un brano elettroacustico.

## 1. AUTOENCODER NSYNTH

Nsynth fa parte di Magenta<sup>1</sup>, un progetto di GoogleBrain<sup>2</sup> per lo sviluppo di strumenti per la generazione di contenuti musicali, immagini, video e testo utilizzando la *deep learning*, attraverso la creazione di una piattaforma *open source* basata sulle librerie TensorFlow<sup>3</sup>.

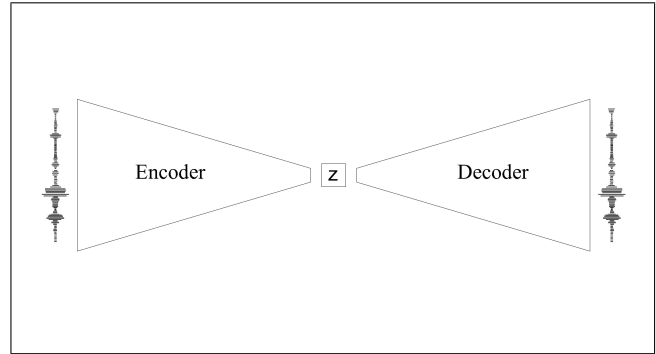
Il sistema Nsynth [1] utilizza un modello di reti neurali per la sintesi del suono ed è un esempio di approccio alla sintesi basato più sulle caratteristiche dei dati d'ingresso (*data-driven-approach*) che sulla struttura dell'algoritmo di elaborazione.

<sup>1</sup> <https://magenta.tensorflow.org>

<sup>2</sup> <https://research.google.com/teams/brain>

<sup>3</sup> <https://www.tensorflow.org>

Copyright: ©2018 Martino Sarolli et Andrea Agostini. This is an open-access article distributed under the terms of the [Creative Commons Attribution License 3.0 Unported](https://creativecommons.org/licenses/by/3.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.



**Figure 1.** Struttura *autoencoder*

Partendo dall'analisi dei dati di *input* il modello è in grado di crearne una rappresentazione interna semanticamente significativa, che può essere utilizzata per controllare ad alto livello variazioni di altezza, durata, timbro e dinamica del suono in uscita.

Il sistema Nsynth è costituito da due elementi principali: un modello e un set di dati per l'apprendimento. Il modello è basato su di un *autoencoder* Wavenet [2], un sistema di codifica e decodifica gestito da due reti neurali multi-livello. Il *dataset* Nsynth è costituito da un ampio insieme di file audio contenenti singole note provenienti da diverse sorgenti strumentali.

### 1.1 Autoencoder

La struttura dell'*autoencoder* comprende tre blocchi: un codificatore (rete neurale per la codifica dei dati d'ingresso), una rappresentazione  $Z$  dei dati codificati e un decodificatore d'uscita (rete neurale per la decodifica).

Il codificatore è un rete neurale multilivello a convoluzione dilatata [3] i cui neuroni implementano una funzione di attivazione non lineare Relu<sup>4</sup>.

I dati audio di input vengono trasformati dal codificatore in rappresentazioni vettoriali  $Z$  a 16 dimensioni e 32 campioni al secondo.

Il decodificatore è una rete neurale multilivello autoregressiva [4] in grado di calcolare il valore di ogni campione in base alle caratteristiche delle rappresentazioni  $Z$  dei dati codificati e ai valori dei campioni precedentemente calcolati, e restituire i dati in formato audio.

<sup>4</sup> dall'inglese *rectified linear unit*, è una funzione di attivazione definita come la parte positiva del suo argomento:  $f(x) = x^+ = \max(0, x)$

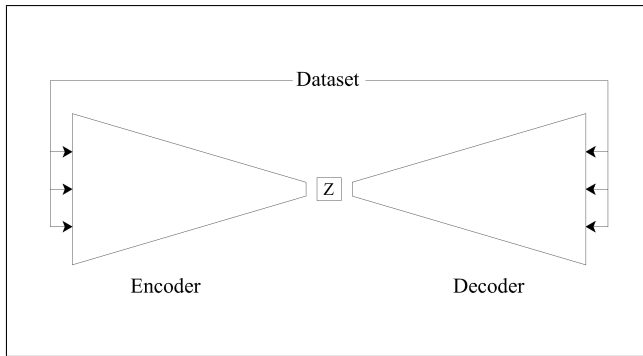


Figure 2. Fase di analisi del *dataset* (*training*)

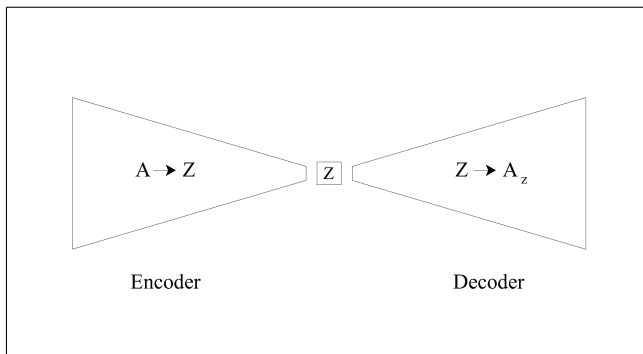


Figure 3. Risultato della fase di *training*

## 1.2 Learning

Il processo di apprendimento del modello è di tipo non supervisionato (*unsupervised learning*) [5] e si divide in tre fasi:

### 1.2.1 Pre-Learning

Comprende la costruzione di un *dataset* per l'addestramento delle reti neurali. Il *dataset* raccoglie l'insieme di dati da cui le reti apprendono le loro rappresentazioni interne. La struttura del *dataset* è determinata dalle caratteristiche dei suoi elementi costitutivi, ed esercita un'influenza decisiva sulle fasi successive del funzionamento del sistema.

### 1.2.2 Training

Durante questa fase le reti di codifica e decodifica analizzano tutti gli elementi del *dataset*: in particolare, la rete di codifica impara a *mappare* da un elemento alla relativa rappresentazione codificata  $Z$  mentre la rete di decodifica impara a *mappare* il percorso inverso, da una rappresentazione codificata  $Z$  ad un elemento del *dataset*.

### 1.2.3 Post-training

Il sistema è ora in grado di ricevere nuovi dati d'ingresso, codificarli in rappresentazioni  $Z$  e decodificarli riportandoli al formato audio di partenza. I risultati dei processi di codifica e decodifica dipendono dalle configurazioni apprese dalle reti durante il *training* e potranno variare in funzione delle caratteristiche dei dati d'ingresso.

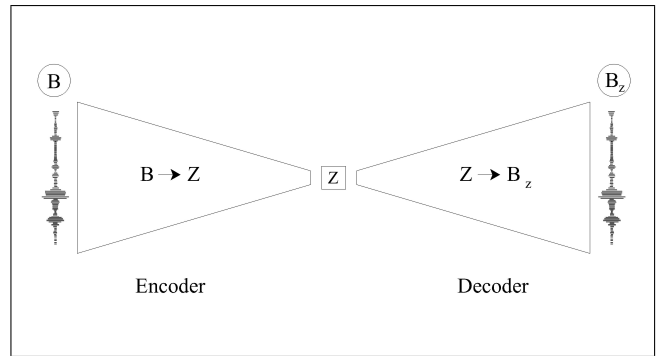


Figure 4. Fase *post-training*, il sistema è pronto per analizzare nuovi esempi

## 1.3 Funzionamento dell'autoencoder

### 1.3.1 Formato audio input

Per diminuire i tempi di calcolo e la potenza computazionale necessaria vengono utilizzati file audio in formato compresso con frequenza di campionamento 16kHz, 8bit di profondità e codifica di compressione  $\mu$ -law<sup>5</sup>.

### 1.3.2 Codifica e rappresentazione dello spazio latente $Z$

Il risultato del processo di codifica è una rappresentazione vettoriale a 16 dimensioni dello spazio latente  $Z$ , con rapporto di compressione 32:1. Si può pensare allo spazio latente  $Z$ , che in linea di principio costituisce uno stadio interno del sistema di auto-encoding, come a una riduzione dell'informazione originale alle sue caratteristiche salienti, dove la salienza è determinata dalla struttura dell'encoder e dal training cui è stata sottoposta. Se però si accede allo spazio latente per manipolarlo è possibile guidare il processo di decodifica: modificando le "informazioni salienti" e immettendole nel decoder si otterranno nuovi materiali audio che le rispecchiano. Le sezioni successive descriveranno come le possibilità di "esplorare" lo spazio latente modificando le rappresentazioni vettoriali  $Z$ , possono essere utilizzate in senso "compositivo" per creare elaborazioni del suono peculiari di questo sistema.

### 1.3.3 Decodifica

Anche senza nessuna trasformazione delle rappresentazioni  $Z$ , lo spettro del suono d'uscita risulta comunque più ricco di parziali rispetto allo spettro iniziale. L'*autoencoder* aggiunge quindi parziali allo spettro dell'*input*. Questo dipende dal fatto che il suono viene ricostruito campione per campione attraverso le funzioni di attivazione non lineari delle reti neurali. La non linearità della propria funzione di attivazione permette alla rete di decodifica di ricostruire in maniera più realistica una gamma maggiore di timbri differenti, aumentando così le capacità espressive del sistema, a scapito della precisione nella ricostruzione della forma d'onda di partenza, che contiene un certo grado di approssimazione. Generalmente l'approssimazione si mantiene

<sup>5</sup> Uno dei principali algoritmi per la codifica e la compressione del suono e della voce, ottimizzato per la telefonia e per la trasmissione via modem e utilizzato nel sistema telefonico americano. La tecnica viene usata dalle applicazioni Java per creare file audio (estensione .au) che siano veloci da trasferire via Internet.

su valori contenuti e vengono prodotte parziali che conservano una relazione diretta con la forma d'onda originaria, arricchendo lo spettro sonoro risultante; occasionalmente si verificano invece evidenti errori di decodifica: la rete "sbaglia" il calcolo di alcuni campioni, se ne "accorge" e corregge rapidamente creando delle repentine discontinuità nella forma d'onda risultante, percepibili all'ascolto come "click".

## 2. DATASET

### 2.1 Struttura del dataset Nsynth

Il *dataset* Nsynth nasce dall'esigenza di disporre di un vasto insieme di esempi audio, costruito sul modello di altri come MNIST [6], CIFAR [7] e Imagenet [8]. Analogamente a questi *dataset*, che sono costituiti da collezioni di immagini singole, Nsynth raccoglie un insieme di file audio contenenti note singole provenienti da diverse librerie commerciali di campioni.

### 2.2 Caratteristiche dataset Nsynth

Il *dataset* comprende 305.979 note musicali della durata di quattro secondi ciascuna, diverse per altezza, timbro ed in-viluppo e provenienti da 1006 differenti strumenti virtuali inclusi in librerie commerciali. La frequenza di campionamento di ogni file audio è di 16 kHz per una risoluzione di 8 bit, l'estensione media per strumento è di 65,4 semitoni e sono rappresentate, in media, 4,75 diverse *velocity* per ogni nota.

Le caratteristiche del *dataset* influenzano i processi di codifica e decodifica sia nel dominio del tempo che della frequenza.

Nsynth dispone di un *dataset* di note singole e di un modello strutturato per analizzare brevi sequenze temporali (divise in *time-frames* da circa 32ms); il risultato è un sistema in grado di operare in maniera più efficiente su file audio non superiori a 4 secondi di durata.

Quando all'interno dell'*input* il modello riscontra leggere variazioni della frequenza fondamentale, poco probabili nel *dataset* da cui ha appreso, aumenterà il grado di approssimazione e nella ricostruzione le fluttuazioni risulteranno accentuate.

Il *dataset* raccoglie prevalentemente suoni ad altezza determinata, di conseguenza l'*autoencoder* rappresenta (codifica) e ricostruisce (decodifica) in maniera più fedele le componenti spettrali armoniche.

Con strumenti ad altezza indeterminata, aumenta la non linearità del sistema e lo spettro del suono in uscita risulta più ricco di parziali inarmoniche non presenti nel suono d'ingresso.

Maggiori sono le dimensioni del *dataset*, maggiore è l'accuratezza con cui le reti codificano e decodificano il suono.

Fornendo al modello un *dataset* costruito appositamente, lo si rende in grado di rappresentare in maniera più efficace le caratteristiche del suono su cui si intende effettuare elaborazioni. Di conseguenza eventuali operazioni sul livello di rappresentazione  $Z$  (stato di codifica) risulteranno più efficaci.

*Dataset* come Nsynth, costituiti da file audio contenenti ciascuno una sola nota per strumento, rendono il modello più adatto ad operazioni sul contenuto spettrale di un suono singolo, piuttosto che su sequenze melodiche, ritmiche o armoniche, per le quali andrebbero costruiti *dataset* appositi. A sequenze di note in ingresso il modello aggiunge una sorta di portamento o glissando dall'altezza dell'ultima nota a quella della nota successiva; in presenza di accordi "salta" da una fondamentale all'altra come se stesse cercando di ricostruire le componenti armoniche di una singola nota.

## 3. CONFRONTO FILTRO AUTOENCODER

E' possibile considerare l'*autoencoder* come un filtro a risposta infinita con funzione di trasferimento variabile, dipendente dalle caratteristiche dell'*input* e dai risultati del processo di apprendimento automatico.

La "funzione di trasferimento" dell'*autoencoder* non è costante e determinata a priori come quella di filtri tradizionali, ma può variare in base alle caratteristiche sonore dell'*input* e del *dataset*, e nel tempo, grazie alle proprie capacità d'apprendimento, migliorare accuratezza e capacità espressive, divenendo in grado di operare su di una varietà sempre maggiore di *input* differenti.

## 4. ESPLORAZIONI DELLO SPAZIO LATENTE

Interessante in senso compositivo è la possibilità di modificare le rappresentazioni latenti  $Z$  (stato di codifica interna) e generare conseguenti elaborazioni del suono risultante.

### 4.1 Caratteristiche delle rappresentazioni $Z$

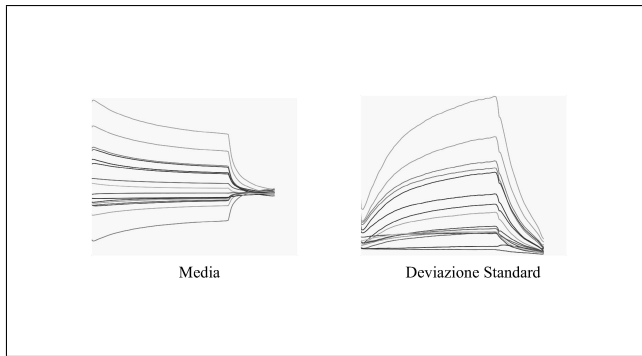
Come uno strumento musicale controlla una vasta gamma di suoni attraverso un numero limitato di parametri, le rappresentazioni compresse sono significative se facilmente manipolabili e in grado di rappresentare una vasta gamma di dati. La particolare struttura del modello "costringe" la rete di codifica a trovare una tale rappresentazione (definita  $Z$ ) forzando il suono attraverso un collo di bottiglia, in modo che l'audio originale a 16kHz sia trasformato in una serie di vettori a 16 dimensioni e circa 30Hz di frequenza di campionamento (32ms per frame).

Le rappresentazioni  $Z$  conservano inviluppo d'ampiezza simile ai suoni originali ed, in alcuni casi, persino informazioni espressive come la presenza di vibrato.

Questo codice latente viene immesso nei *layer* della rete di decodifica al momento della generazione dell'*output* e funziona come un'eccitazione/attivazione di un sistema di *feedback*<sup>6</sup> non lineare<sup>7</sup>; in caso di sua assenza, la rete impara a restare "silenziosa". Considerato in questa prospettiva, il *decoder* si può paragonare ad un filtro a risposta impulsiva infinita non lineare che riceve in *input* le rappresentazioni latenti.

<sup>6</sup> il decodificatore è una rete autoregressiva v.sezione 1.1

<sup>7</sup> i neuroni della rete di codifica implementano una funzione non lineare Relu v.sezione 1.1.



**Figure 5.** Rappresentazioni  $Z$  Media e Deviazione Standard

#### 4.2 Media e deviazione standard

Le rappresentazione  $Z$  di media e deviazione standard del dataset di training, considerato nella sua interezza, mostrano quali caratteristiche accomunano e quali differenziano tra loro i suoni all'interno del *dataset*. La media ha un attacco rapido e un lento decadimento seguito da un altro decadimento repentino alla fine della nota. Relativamente all'involuppo d'ampiezza dei suoni, questo tipo di andamento si riscontra quindi molto frequentemente all'interno del dataset.

La deviazione standard aumenta invece nella fase centrale, dove si concentrano evidentemente le differenze maggiori tra i suoni del *dataset*. Generalmente nei momenti in cui il modello non riesce a riprodurre le caratteristiche di un suono d'ingresso riproduce quelle della media degli altri suoni analizzati.

#### 4.3 Scaling

Moltiplicando le rappresentazioni  $Z$  per un coefficiente scalare si ottengono i seguenti risultati: con coefficienti minori di 1, il *decoder* non riesce a mantenere un suono prolungato consistente.

Moltiplicando per valori maggiori di 1, aumenta progressivamente la distorsione armonica. Superata una certa soglia, il sistema va in saturazione e oscilla continuamente tra i valori estremi.

#### 4.4 Scaling e media

Modificando il rapporto tra la rappresentazione  $Z$  di un suono con la rappresentazione  $Z$  della media del dataset, si possono isolare e riscaldare solo le caratteristiche di un suono che differiscono dalle caratteristiche medie di tutti i suoni del dataset tramite la formula seguente:

$$Z_{output} = Z_{media} + Scaling * (Z_{input} - Z_{media})$$

Sottraendo alla rappresentazione  $Z_{input}$  di un suono di input quella  $Z_{media}$  della media del dataset si ottiene una rappresentazione  $Z$  delle caratteristiche sonore dell'input che differiscono maggiormente da quelle della media di tutti i suoni del dataset.

Riscaldando ( $Z_{input} - Z_{media}$ ) per un valore compreso tra 0 ed 1 si possono attenuare le caratteristiche che rappresenta, che vengono diversamente accentuate se il fattore di scaling è maggiore di 1.

Sommando il risultato dello scaling a  $Z_{media}$  si ottiene una  $Z_{output}$  in cui le caratteristiche peculiari della  $Z_{input}$  rispetto a  $Z_{media}$  risultano accentuate oppure attenuate.

In questo modo è quindi possibile accentuare o attenuare le caratteristiche proprie di un determinato suono rispetto a quelle della media di tutti i suoni analizzati, ottenendo quelle che, utilizzando una metafora visuale, potrebbero assomigliare a diverse "caricature" del suono di partenza.

#### 4.5 Interpolazioni

Attraverso operazioni di calcolo vettoriale sulle rappresentazioni  $Z$  si ottengono particolari trasformazioni dei materiali sonori, che possiedono all'ascolto un carattere distintivo, peculiare di questo sistema di elaborazione.

#### 4.6 Media di rappresentazioni $Z$

La possibilità di calcolare la media tra due (o più) rappresentazioni  $Z$  permette di combinare insieme le caratteristiche timbriche e dinamiche di due (o più) suoni diversi, ottenendo un effetto differente rispetto al mixaggio tradizionale. Modificando i rapporti di ampiezza tra i suoni d'ingresso si ottengono variazioni della loro combinazione risultante: le componenti provenienti dal suono di ampiezza maggiore avranno maggiore influenza sulle caratteristiche generali del timbro e viceversa.

#### 4.7 Crossfade

La variazione dinamica dei rapporti d'ampiezza (*cross-fade*) produce una variazione dinamica dell'interpolazione nel tempo tra due timbri. Anche in questo caso, si ottiene un risultato differente rispetto ad un normale *cross-fade* dell'ampiezza di due suoni.

#### 4.8 Vector analogies

Utilizzando modelli generativi è possibile trovare analogie vettoriali (*vector analogies*) [9], regolarità dello spazio latente che permettono l'elaborazione di aspetti specifici dell'*input* e la generazione di nuovi esempi estranei al *dataset*. Nel caso di modelli per la generazione di immagini e testo, i risultati sono facilmente interpretabili e semanticamente riconoscibili.

Due esempi per illustrare meglio il concetto:

Nello spazio latente creato da alcuni modelli linguistici la rappresentazione vettoriale di "Re - Uomo + Donna" è un vettore molto vicino a quello di "Regina".

Considerando un ipotetico modello che abbia analizzato soltanto le parole "Re", "Uomo" e "Donna", e si desideri ottenere da lui un esempio della parola "Regina" (un elemento quindi non presente nel *dataset* di partenza), il modello sarà in grado di ricostruire una rappresentazione semanticamente molto vicina alla parola "Regina" attraverso l'operazione di calcolo vettoriale seguente:

RE : UOMO :: DONNA : ?



## 5. ALGORITMI PER L'ELABORAZIONE SONORA

In questa sezione le principali tecniche di elaborazione dell'*autoencoder* di Nsynth verranno illustrate attraverso alcuni esempi sonori tratti dalla composizione acusmatica "Parafulmini e parassiti" di Martino Sarolli. Per ognuna delle tecniche di trasformazione applicate al materiale sonoro del brano verrà descritto il processo di realizzazione e riportato il relativo codice di programmazione (in linguaggio Python, versione 3.6).

### 5.1 Requisiti di installazione Nsynth

Come indicato all'inizio del capitolo primo, il sistema Nsynth fa parte del progetto Google Magenta e il requisito principale per il suo funzionamento è l'installazione del pacchetto Magenta <sup>8</sup>.

Per poter utilizzare il codice presentato è necessario installare anche i pacchetti os, numpy e IPython e importarne i seguenti moduli e funzioni:

```
import os
import numpy as np
from magenta.models.nsynth import utils
from magenta.models.nsynth.wavenet import fastgen
from IPython.display import Audio
```

La frequenza di campionamento dei file audio utilizzati dall'*autoencoder* dev'essere pari a 16khz.

### 5.2 Codifica

Il passo successivo per l'elaborazione di un file audio consiste nel crearne una rappresentazione nello spazio latente *Z* dell'*autoencoder*.

Le trasformazioni del materiale sonoro avvengono a livello di rappresentazione *Z*.

La funzione `fastgen.encode` <sup>9</sup> codifica uno o più file audio a 16Khz creando le relative rappresentazioni *Z* (qui definite *encoding* dei file) <sup>10</sup>.

```
fastgen.encode(wav_data, checkpoint_path,
               sample_length=64000)
Generate an array of embeddings from an
array of audio.
Args:
  wav_data: Numpy array [batch_size,
                        sample_length]
  checkpoint_path: Location of the
                  pretrained model.
  sample_length: The total length of the
                 final wave file, padded with 0s.
Returns:
  encoding: a [mb, 125, 16] encoding (
            for 64000 sample file).
```

La funzione restituisce *encoding* da 32 campioni al secondo per 16 dimensioni.

<sup>8</sup> <https://github.com/tensorflow/magenta>

<sup>9</sup> Il modello richiesto come secondo argomento è disponibile al seguente link: <http://download.magenta.tensorflow.org/models/nsynth/wavenet-ckpt.tar> - alcuni degli esempi di codice presentati nella sezione 5 sono disponibili al seguente link : <https://magenta.tensorflow.org/nsynth-fastgen>

<sup>10</sup> <https://github.com/tensorflow/magenta-demos/blob/master/jupyter-notebooks/NSynth.ipynb>

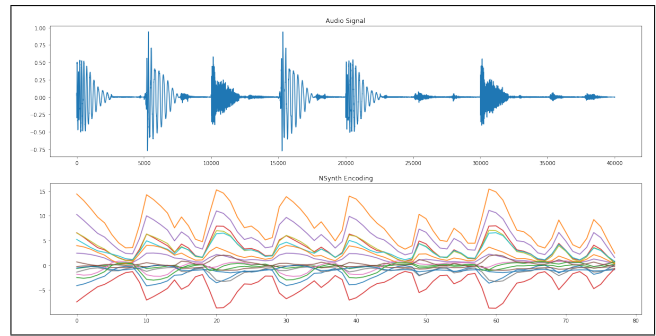


Figure 8. Confronto tra file audio e relativo *encoding*

Visualizzando graficamente file audio e relativo *encoding* è possibile notare come, in seguito al processo di codifica, una certa correlazione con l'andamento delle variazioni d'ampiezza del file audio venga mantenuta nella rappresentazione *Z*.

### 5.3 Decodifica

La funzione `fastgen.synthesize` [11] accetta in ingresso un *encoding* che decodifica e riporta in formato audio a 16Khz.

```
fastgen.synthesize(encodings, save_paths,
                  checkpoint_path='model.ckpt-200000',
                  samples_per_save=1000)

Synthesize audio from an array of
embeddings.
Args:
  encodings: Numpy array with shape [
            batch_size, time, dim].
  save_paths: Iterable of output file
             names.
  checkpoint_path: Location of the
                  pretrained model.
  samples_per_save: Save files after
                   every amount of generated samples
.
```

Senza operare nessuna trasformazione sull'*encoding*, il processo di risintesi delle reti neurali dell'*autoencoder* non restituisce comunque il suono allo stato originario, ma modifica lo spettro di partenza alterando i rapporti di ampiezza tra le parziali e aggiungendo nuove componenti armoniche e inarmoniche.

I risultati del processo di decodifica variano molto in funzione delle caratteristiche del suono codificato.

Nella composizione del brano, per evocare effetti di trasparenza e opacità questa tecnica è stata utilizzata su suoni singoli nel registro grave (fiati e percussioni), su accordi di singoli strumenti (organo) e frammenti di polifonie vocali.

Con suoni singoli nel registro grave, la decodifica produce una sorta di filtraggio non lineare delle componenti acute, e, in senso metaforico, maggiore "opacità" nella riproduzione del suono.

Nei casi in cui sono presenti di più note fondamentali simultanee, la rete di decodifica, avendo appreso a ricostruire soltanto suoni monofonici, si trovano senza punti di riferimento ma cercano comunque di ricostruire il suono saltando rapidamente da una nota fondamentale dell'insieme polifonico ad un'altra. Il comportamento non lineare della rete altera sia le articolazioni temporali che le relazioni

dinamiche tra le componenti originarie dei frammenti polifonici, restituendo un risultato sonoro timbricamente differente che conserva tuttavia caratteristiche armoniche affini al materiale di partenza. Questo processo è stato utilizzato nel brano per generare variazioni timbriche e dinamiche all'interno di un contesto armonico statico.

#### 5.4 Saturazione non lineare

Moltiplicando un *encoding* per un coefficiente compreso tra 1 e 2 si ottiene un particolare effetto di distorsione non lineare anch'esso variabile in funzione del tipo di suono codificato. Nella composizione sono stati elaborati con questa tecnica suoni monofonici vocali nel registro acuto.

```
|| encoding_sat = encoding * 1.4
|| fastgen.synthesize(encoding_sat, save_paths=
||   sat.wav')
```

#### 5.5 Interpolazioni

Per creare concatenamenti timbrici di suoni eterogenei si ricorre alla possibilità di sommare tra loro *encoding* di file audio differenti e calcolarne la media risultante.

```
|| encoding_interp = (encoding1 + encoding2 +
||   encoding3) / 3.0
```

Decodificando l'*encoding* risultante si ottiene la combinazione sonora.

```
|| fastgen.synthesize(encoding_interp, save_paths
||   ='interp.wav')
```

Il timbro risultante è peculiare del sistema di elaborazione degli *encoding* e non ha corrispondenze con altre tecniche di combinazione sonora come la cross-synthesis [12] e il mixaggio "tradizionale". Il concatenamento interessa sia l'involuppo d'ampiezza che quello spettrale. Nel dominio del tempo si possono ottenere risultati interessanti interpolando suoni con attacco e decadimento rapido con altri caratterizzati da fasi di *sustain* prolungate. Calcolando la media di suoni con la stessa fondamentale e timbro differente si ottiene una particolare combinazione inarmonica delle parziali dei diversi spettri. Durante la realizzazione del brano sono state sperimentate diverse concatenazioni realizzate tramite le tecniche di interpolazione descritte sopra, tra cui ad esempio voci nel registro acuto e flauti, corni e voci nel registro grave, percussioni e versi di animali.

#### 5.6 Crossfade

Applicando agli *encoding* tremoli e dissolvenze si possono creare interpolazioni dinamiche tra timbri diversi. La funzione *fade* applica una dissolvenza in ingresso ad un primo *encoding* ed una in uscita ad un secondo.

```
|| def fade(encoding, mode='in'):
||   length = encoding.shape[1]
||   fadein = (0.5 * (1.0 - np.cos(3.1415 * np.
||     arange(length) / float(length))))
||     reshape(1, -1, 1)
||   if mode == 'in':
||     return fadein * encoding
||   else:
||     return (1.0 - fadein) * encoding
```

La somma dei due *encoding* genera un *crossfade* timbrico simile ad una sorta di *morphing* tra due suoni.

```
|| def crossfade(encoding1, encoding2):
||   return fade(encoding1, 'out') + fade(
||     encoding2, 'in')
|| fastgen.synthesize(crossfade(enc1, enc2),
||   save_paths=['crossfade.wav'])
```

## 6. CONCLUSIONI E SVILUPPI

Il presente studio nasce da un processo graduale di scoperta delle diverse possibilità offerte dai sistemi di apprendimento automatico applicati alla pratica musicale sperimentale. Tra quelli considerati, gli algoritmi basati sulle reti neurali si sono rivelati affini al tipo di ricerca musicale che si intende perseguire: una prassi compositiva diretta all'esplorazione di concetti di coesistenza, multilinearità e stratificazione di componenti eterogenee e velocità differenti.

L'espressività esibita dalle reti nei loro comportamenti non lineari di decostruzione e ricostruzione del materiale sonoro, guidati dalla mano del compositore che vi interagisce attraverso il controllo del loro bagaglio di conoscenza immesso in forma di dati e la manipolazione delle loro rappresentazioni interne hanno permesso di ottenere risultati inaspettati e forieri di ulteriori opportunità di esplorazione.

Attraverso l'indagine sulle caratteristiche dello spazio latente interno alle reti è possibile sperimentare relazioni inusitate tra elementi sonori senza un legame apparente e indirizzare il processo compositivo verso nuove territorialità.

In futuro si cercherà di disegnare una mappa di questo nuovo spazio sonoro attraverso l'impiego di algoritmi per la visualizzazione di dati multidimensionali (come la tecnica T-SNE [13]); l'esplorazione delle potenzialità del modello proseguirà partendo dall'applicazione di algoritmi normalmente utilizzati per differenti tipologie di dati, come librerie per l'*image-processing*, all'audio in forma codificata dalle reti, per ottenere nuove elaborazioni delle dimensioni temporali e spettrali del suono.

La disponibilità di un'ampia e strutturata collezione di dati audio come il *dataset* Nsynth permette di poter "allenare" anche modelli differenti rispetto all'autoencoder come, ad esempio, una coppia di reti GAN (*generative adversarial networks*)[14][15] ed ottenere ulteriori nuove generazioni di materiali sonori.

## 7. RICONOSCIMENTI

Il presente lavoro costituisce un estratto della tesi per il Biennio di Musica Elettronica del Conservatorio G. Verdi di Torino, discussa da Martino Sarolli con relatore Andrea Agostini, dal titolo: "Sistemi Deep Learning per l'Elaborazione Digitale Del Suono: possibilità di integrazione nella prassi compositiva elettroacustica".

## 8. BIBLIOGRAFIA

- [1] J. Engel, C. Resnick, A. Roberts, S. Dieleman, D. Eck, K. Simonyan, and M. Norouzi, "Neural Audio Synthesis of Musical Notes with WaveNet Autoencoders," *ArXiv e-prints*, Apr. 2017.
- [2] A. van den Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, and K. Kavukcuoglu, "WaveNet: A Generative Model for Raw Audio," *ArXiv e-prints*, Sept. 2016.
- [3] Y. LeCun, K. Kavukcuoglu, and C. Farabet, "Convolutional networks and applications in vision.," *Circuits and Systems (ISCAS), Proceedings of 2010 IEEE International Symposium on*, 2010.
- [4] K. Gregor, I. Danihelka, A. Mnih, C. Blundell, and D. Wierstra, "Deep AutoRegressive Networks," *ArXiv e-prints*, Oct. 2013.
- [5] X. Wang and A. Gupta, "Unsupervised learning of visual representations using videos," *CoRR*, vol. abs/1505.00687, 2015.
- [6] Y. LeCun, C. Cortes, and C. Burges, "The mnist database of handwritten digits," 1998.
- [7] A. Krizhevsky and G. Hinton, "Learning multiple layers of features from tiny images," 2009.
- [8] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A Large-Scale Hierarchical Image Database," *CVPR09*, 2009.
- [9] T. White, "Sampling Generative Networks," *ArXiv e-prints*, Sept. 2016.
- [10] O. Levy, Y. Goldberg, and I. Ramat-Gan, "Linguistic regularities in sparse and explicit word representations.," in *CoNLL*, pp. 171–180, 2014.
- [11] P. Ramachandran, T. Le Paine, P. Khorrami, M. Babaeizadeh, S. Chang, Y. Zhang, M. A. Hasegawa-Johnson, R. H. Campbell, and T. S. Huang, "Fast Generation for Convolutional Autoregressive Models," *ArXiv e-prints*, Apr. 2017.
- [12] C. Donahue, T. Erbe, and M. Puckette, "Extended convolution techniques for cross-synthesis," in *Proceedings of the International Computer Music Conference*, 2016.
- [13] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *Journal of Machine Learning Research*, vol. 9, pp. 2579–2605, 2008.
- [14] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative Adversarial Networks," *ArXiv e-prints*, June 2014.
- [15] C. Donahue, J. McAuley, and M. Puckette, "Synthesizing Audio with Generative Adversarial Networks," *ArXiv e-prints*, Feb. 2018.

# ANALYSIS OF SOUND DESIGN PRACTICES [ASDP]. RESEARCH METHODOLOGY

**Laura Zattra**

Conservatorio di Musica di  
Rovigo  
STMS – IRCAM, CNRS  
laura\_zattra@yahoo.it

**Nicolas Misdariis**

STMS  
Ircam-CNRS-SU,  
Sorbonne Université  
misdariis@ircam.fr

**Frank Pecquet**

Pantheon Sorbonne  
University – CNRS  
fpecquet@univ-  
paris1.fr

**Nicolas Donin**

STMS Ircam-CNRS-  
SU, Sorbonne  
Université  
donin@ircam.fr

**David Fierro**

CICM EA 1572  
MUSIDANSE  
EDESTA - Paris8  
davidfierro@gmail.com

## ABSTRACT

This contribution describes an on-going project started in March 2018 entitled “Analysis of Sound Design Practices” [ASDP]. The team comprises members from the LABEX CAP (Création, Arts et Patrimoines); the ACTE Institute (Arts Creation Theory Aesthetics); IRCAM’s APM (Analysis of Musical Practices) and PDS (Sound Perception and Design) teams.

By means of a large-scale study – based on web documentation, an online questionnaire, interviews and analytical results – the team is working towards understanding the creative process as well as the identity of Sound Design and Sound Designers in particular, in Europe. The main target concerns the development of a consistent sociological, geographical and historical knowledge of the discipline, and, consequently, to better identify the relevant issues in both artistic, technical or scientific commitment.

The purposes of this article is to describe (a) the general framework of the project; (b) the methodology of the ongoing study (collaborative tools, writing and sending a questionnaire); (c) initial developments and results - terminology and definition, translation issues, conception of a database, listing of professionals and different institutions in Europe, first feedback analysis.

## 1. INTRODUCTION

The pursuit of designing sounds in human activities appears to be as ancient as the human race: using automatic alarm signals and striking clocks in ancient Greece, combining a clepsydra with a water organ ([1] cited in [2]); designing the acoustics of ancient Greek and Roman theatres [3] [4]; building bell towers and using bells as earcons [30] since their introduction, according to tradition, by Paulinus, bishop of Nola - a town of Campania, Italy - during the V century AD [5]; using sound effects to evoke emotions and underscore actions during the Commedia dell’Arte performances from the XVI through the XIX century [pp.97-98: 6];

*Copyright: © 2018 Zattra et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution License 3.0 Unported](https://creativecommons.org/licenses/by/3.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.*

up to inhuman torture practices using sounds in ancient Chinese times [p.109: 7]. If we except the previous example, such activities are at the intersection of different disciplines: art, craftsmanship [35] [36], industry (at large) and science.

Professional practices relating to sound design come from the depths of time and they were recognized as such only on the last century. These activities are carried out in different fields: architecture and urban environment, visual and digital arts, cinema and TV, advertising, ecology and acoustic regulation, industry, communication and marketing. Such professional practices are therefore characterized by interdisciplinarity, as expressed by Murray Schafer in 1977 [10]. Sound designer must have acoustical knowledge, sound processing skills, psychological, information technology, cognitive, socio-economic know-how, musical and aesthetical talents.

In many areas, sound designers have already a recognized career path. In theatre e.g., the process of designing/creating sound and ambience has been described as earlier as 1936 by Frank Napier [8] [31]. At the same time one might consider as well Vitruve architectural conception of early theatres (Antiquity) as pertinent sound design practices, calling for specific fields of sound research/creation such as acoustics, architecture and drama (voice and noise production).

However, despite the long history of such activity and the related jobs involving sound design, the terminology of “sound design” and “sound designer” have remained puzzling and, as such, quite new and not yet clearly accepted as a specific profession in different places in Europe. As such, associating “sound” with “design” was first mentioned by David Collison (1959) [4] in theatre to describe together with stage manager audio engineering activity. In the film industry, years later, this terms appear together first in the credits of the movie *Apocalypse Now* (1979), when, with the introduction of new audio technology, Walter Murch (followed by, or at the same time of, Ben Burt) was able to «collapse for the first time the duties of sound recordist, editor, and mixer to experiment with sound in unprecedented ways» [p.288: 9]. The discipline has also been explored and theorized by Raymond Murray Schafer during the same period, although in a totally different perspective [10].

As a field of science, research and development, sound design leads to a growing interest in academia and concerns various domains, whether it be industrial (economic) or artistic, in which innovative engineering meets artistic creation. Although sound design studies and research are caught into different fields of practices - earlier referred to as human and applied sciences, there need specific studies in order to connect sound design practices as a subdomain of the science of design.

Moreover, one may consider that sound design still concerns related domains as distinct types of vocation. As a matter of fact, on Internet, everyone may call him/herself a sound designer, although practicing different activities, when it comes to work on sounds. Being a sound designer might be even 'trendy'. The impression is that the profession of sound designer is severely compromised by the general attitude reported in the quotation by Mark Kilborn.

First, if you're seeking sound design roles, call yourself a Sound Designer. Don't obsess about your lack of shipped titles or experience. For the first year or two I was trying to get my foot in the door, I obsessed over whether it was disingenuous and/or presumptuous to call myself what I so desperately wanted to be. It was silly. If you design sound, you are a sound designer. Be confident about that [12].

Either the concept defining this word is self-evident, and does not need to be explained, or there is actually something that all the different interpretations of the term have in common, and must be clarified. We believe that Mark Kilborn's words, similarly to many others expressing doubt on the meaning of the term, bring forward to a necessary research on the topic. In other words, the question could be whether or not there is a real difference between a designer and a sound designer (beyond the specific skills). Can a designer create sound? «Sound design should not try to find any definition of itself other than within design», sound designer Louis Dandrel wrote. «There ought to be a definition of design with sound; why should we remove from sound design one of its sensory attributes?», he added (cited in [37]). Today one may simply recognize that design concerns better quality [11]. Sound has increased power in society and sound design is a matter of improvement, whether it be about environment, objects, services or communication [32].

The ASDP project is aimed at covering the serious gap in sound design research, that is to say how sound design is carried out (process and methods), who is the sound designer (practices and behaviours), what is or can be considered a sound design (artefact) [14] [17]. By adopting an approach globally inspired by design research [13] [14] (see Section 2), including socio-economic considerations of sound design practices, the ASDP project focuses on the main actors of the discipline, the sound designers, together with an investigation on their activities and according to

pertinent design contexts such as artistic, musicological, sociocultural, technical/technological or ecological domains.

By means of large-scale studies – realized through an online questionnaire and interviews – the project has different objectives: promoting the understanding of creative processes underlying sound design practices; building a consistent (historical, anthropological and methodological) knowledge about the discipline; and consequently, better evaluating how sound design impacts the artistic, technical or scientific domains.

The following sections describe the methodology, research protocol and the first phases of the project (the questionnaire), as well as the hypotheses and initial thoughts on our research topic. This article is meant to be a vehicle «to think [our] way through the research process» [p.2: 15], because our «research is not a straightforward process» (ibidem). Writing about our methodological choices and our research protocol helps us establishing a more coherent framework to our future outcomes. By sharing the protocol of this investigation, a collaborative work with a defined method, we also hope that such a learning process will bring back self-evaluation.

## 2. THEORETICAL FRAMEWORK AND PREVIOUS RESEARCH

The project is inspired by different design research approaches elaborated by [16] [14] and more recently [11]. From that point of view, sound design may be considered as a proper «discipline», a research object «on its own», and a specific, or «third» culture being halfway «between arts and humanities» [16]. The transposition of original paradigms established by Cross to the field of science of design [14] leads us to propose a framework for acquiring knowledge about the discipline. Inspired by Cross' framework, there should be three distinct researches 'loci' for sound design : (1) «people» (activities): status and practices of the sound designers [17]; (2) «processes» (methods): status of the sound design itself, innovative methods and tools with regards to sound prototypes (mockups, sketches, intermediary objects), creativity/fixation mechanisms; (3) «products» (artefacts): status of the designed sounds, new forms, formats, listening situations or author/listener relationships [pp.125-6: 14] [18]. The ASDP project explores one of these three sources of sound design knowledge: «people» [14].

Literature dedicated to sound design is important, and has continued to grow especially over the last three decades. One of the preliminary goals was to construct a theoretical framework based upon the bibliographical references devoted to historical, practical and epistemological issues: examples include among others [7] [19] [20] [21] [4] [22] [23] and dozens of articles and books collected over the years. This research will continue all along our project.

Moreover, the project is also (but not only) based upon literature dedicated to Sonic Interaction Design (SID), an emerging field at the intersection of interaction design, sound and music computing, auditory display, sonic arts and acoustics. For around 15 years now, SID researchers have been working actively to formalize tools, methodologies and conceptual frameworks in this scientific domain. A breakthrough advance to auditory interfaces and displays research has been the project SOb (the Sounding Object 2001-2003, IST-2000-25287 [38]) coordinated by Davide Rocchesso, followed among others by the COST Action on Sonic Interaction Design (2007-2011, ICT COST Action IC0601 [39]) and the SkAT-VG project (Sketching Audio Technologies using Vocalizations and Gestures) [40]. Those research and publishing projects (see also [41] and [42]) have helped a community of scientists and practitioners to federate.

Our project team is formed by researchers who have been working on Design, Sound Design, Semiotic of Art and Design, Collaborative Musical Creation, Music Composition, Electroacoustic Music Studies, Information Technology and Electronic Engineering.<sup>1</sup> This projects derives from previous studies on the «Designerly way of thinking the sound design concepts» [18], the «Collaborative environment in computer music composition» [24] and the «Meaning of Sound Design - main concepts and creative approaches [25].

### 3. TERMINOLOGY AND TRANSLATIONS

#### 3.1 Qualifying Sound Design

Sound design can be considered as a field of design which consists in taking into account the *sonic dimension* of an object (matter and form) within the *designing process* - any sounding features at this stage – the term ‘object’ being able to be embedded either in the *tangible* (manufactured products), *digital* (man-machine interface) or *spatial* (environmental context) dimension. Sound Design can be seen as:

- a collective practice [26],
- reassessing the fundamentals of design practice with sound matter, answering to multiple requests (specifications),
- in case being constrained by composition rules (translation of the specifications in sonic dimensions),

- targeting a goal grounded on functional and/or aesthetic criteria (the search of the ‘best’ or the ‘beautiful’ sound),
- a practice where one’s identity (musical/artistic) is offered in the service of the project (non authoriality);
- where the (personal) search of beauty runs parallel with taking into consideration and satisfying production constraints,
- on several occasion, grounding on scientific results from disciplines such as psychoacoustics, psychology, etc. (scientific attitude).

In other words, sound design can be considered as a multidisciplinary creation, involving specific means of production, based on user targeted data, combining both function and aesthetics. When all these skills/competences/factors are gathered, although in related proportions, we may call it sound design. The latter is therefore dependent upon transversal competences that do converge all together to sound design expertise. While sound design practices are distinct and varied due to many domains of sound application, the items listed above defines its working activity.

Such a conception results from a comparative analysis of different materials : published literature, books and scientific papers, blogs, web articles, landing pages of training and master courses on sound design,<sup>2</sup> our past experiences and researches in the audio-visual production or film industry (Film, TV, Video game industry, theatre, scenography, Foley-bruitage); survey on product and services (transportation industry, medical industry, domestic electrical industry, digital industry such as auditory display / human-machine interfaces, sonic interactive devices); studies on places / spaces (architecture, soundscape, acoustic regulation, museography); communication (marketing, sound branding, radio broadcasting); and at last performing / fine arts (music, instrument making, sound art).

#### 3.2 The Sound Designer Activity

If the sound designer professional activity links to various categories of design - space, message, product, it’s always within design competences in research (innovation), creation (imagination) and communication (industrial and marketing strategies). A non exhaustive list of sectors with related sound design practices includes: sound illustration, music composition, audio post-production, soundscaping, sound branding, acoustic and psychoacoustic experiments, audio ergonomic trends, health and ecology, engineering and signal processing, sonification and signaletic, industrial design and marketing. Our team wrote a questionnaire sent last Summer - July 2018 (section 5), to inventory sound design practices with regard to existing professional activities to highlight the social, cultural

<sup>1</sup> Our research team is working on behalf of LABEX CAP (Laboratory of Excellence « Creation, Arts and Heritage », <http://labexcap.fr/labex-cap/>) in Paris, an observatory and experimental research laboratory working under the auspices of the French Ministry of Culture as well as in collaboration with other French institutions: the ACTE Institute - Arts, Créations, Théories, Esthétiques - in Paris (<http://www.institut-acte.cnrs.fr/semiotics/>); the IRCAM/APM team « Analyse des Pratiques Musicales » (<https://www.ircam.fr/recherche/equipes-recherche/apm/>); the IRCAM/PDS team « Sound Perception and Design » (<https://www.ircam.fr/recherche/equipes-recherche/pds/>).

<sup>2</sup> Such as the Master option ‘mention design sonore’ des écoles ESBA, ENSCI, IRCAM lemans.esba-talm.fr/études/option-design-sonore/.

and economic profile of a sound designer, seeking different types of information in education, methods, processes, project managing, marketing issues together with scientific and artistic researches. The questionnaire aims at defining a larger profile of sound design activities throughout new sound design practices.

### 3.3 Translating Sound Design

Nevertheless, at this stage of the research, the project’s short-term perspective is first to find the ‘practitioners’ in order to better qualify sound design practices. Accordingly, amongst professionals that we already know from our previous inquiries, mapping sound designers leads to searching for names in Internet. While English is the obvious language of this research project (the mother tongue, and etymologically the original terms to define this activity), we systematically sought a proper translation for the words “sound designer” in every single country being part of the European Community, since no compromise so far has been clearly made to qualify this. Such a strategy in research remains necessary to refine the investigation, integrating cultural differences on the topic and/or the profession, and applying the right keywords for search engines. Translations have been obtained through online translation software (context.reverso, linguee) and bilingual dictionaries. Translated results on sound design(er) were coupled with keywords such university, agency, master, course, project, profession, etc. Table 1 represents the different translations of the term Sound Design in the EU member states.

| EUROPEAN COMMUNITY 2018 | Sound Design   | Sound Designer  |
|-------------------------|--|---|
| GERMANY                 | Sound Design, Audiogestaltung                            | Sounddesigner - Klangarchitekt - Klanggärtner                           |
| AUSTRIA                 | Sound Design, Audiogestaltung                            | “ “ “   |
| BELGIUM                 | cf. France   | cf. France  |
| BULGARIA                | звук дизайн  | звук дизайнер   |
| CYPRUS                  | cf. grècque / turc                                       | cf. grècque / turc  |
| CROATIA                 | dizajn zvuka   | dizajner zvuka  |
| DENMARK                 | lyddesign  | lyddesigner   |
| SPAIN                   | Diseño sonoro, Diseño de Sonido                          | Diseñador de sonido, Sound designer                                     |
| ESTONIA                 | heli disain  | heli disainer   |
| FINLAND                 | äänisuunnittelu  | äänisuunnittelija   |
| FRANCE                  | Design sonore  | Designer sonore - Synonymes: concepteur sonore, Illustrateur sonore     |
| GREECE                  | σχεδιασμό ήχου   | σχεδιαστής ήχου   |
| HUNGARY                 | hang design  | hangtervező   |
| IRELAND                 | dearadh fuaim  | dearthóir fuaim   |
| ICELAND                 |  |   |
| ITALY                   | Sound design, Progettazione del Suono, Disegno del suono | Sound designer - Synonyme: Progettista del Suono, disegnatore del suono |
| LATVIA                  | skaņas dizains   | skaņu dizaineris  |
| LITHUANIA               | garso dizainas   | garso dizaineris  |
| LUXEMBOURG              |  |   |
| MALTA                   | disinn tal-hoss  | disinjatur tal-hoss   |

|                |                             |   |
|----------------|-----------------------------|---|
| NETHERLANDS    | Geluidsonwerp               | Geluidsonwerper                                   |
| POLAND         | Sound design                | Sound designer, operator dźwięku                  |
| PORTUGAL       | Sound design, Design do som | Sound designer, Criador de efeitos sonoros        |
| ROMANIA        | Sound design                | Sound designer, Designer audio, Designer de sunet |
| UNITED KINGDOM | sound design                | sound designer                                    |
| SLOVENIA       | zvočna zasnova              | oblikovalec zvoka                                 |
| SLOVAKIA       | zvukový dizajn              | zvukový dizajner                                  |
| SWEDEN         | ljuddesign                  | ljuddesigner                                      |
| CZECH REPUBLIC | zvukový design              | Zvukový designér                                  |
| SERBIA         | звучни дизајн               | звучни дизајнер                                   |

**Table 1.** Translation of the term Sound Design in the European Unity Member States (including UK).

At the same time, we have developed a terminological analysis of the term “sound design”. In France for instance, the term *design* comes first before that of sound (‘design sonore’, ‘designer sonore’). According to us, there is an apparently slight, but nevertheless eloquent, distinction with the English version, where the term sound comes first before that of design [32]. If the grammar has obviously an impact on this lexicon, it also affects epistemology. The use of an English expression (Sound Design) does not mean that it determines logically the exact same knowledge (and/or practices) in different countries, in English, according to the common sense, a sound designer is working for the film industry [34]. Otherwise in France, science and sound design have been associated only recently [32, 44]. Such an expression implies skills and knowledge in sciences of sound (computer processing, signal processing, acoustics,...) and refers to the science of design and sound, the “science of applied sound”. This type of nuances is therefore necessary for epistemological reason and helps defining sound design practices.

Although rare in Italian, the expression *Progettista del suono* [33] or *disegnatore del suono* emphasizes on design methods such as planning, conceptualizing and engineering [27], which inevitably are etymologically at the foundation of design practices.

## 4. PROJECT CONTENT AND METHODOLOGY

During the first weeks of this project (starting date: March 2018), we used an online collaborative platform (wiki and googledocs) to better control the collected data - names of sound designers, institutions, websites, and bibliographical references. Together with this collective work such methods as brainstorming, shared communication - meetings, email, phone calls, skype calls and informal report writing, complete this investigation. While using these methods the team shares a common belief - not to mention a specific design thinking process, where technology provides an environment that embodies “social-constructivist

principles” [29] and where each member can review and update contents in a progressive co-construction knowledge.

**4.1 Database: Mapping the profession of Sound Design**

This section shows the results of our cartography. So far we have collected a list of more than 150 institutions - agencies, universities or academies, related to sound design production and sound design training in Europe. This survey aims at building a more precise cartography of *places* according to sound design approaches, creation, training and engineering.

This cartography shows a collection of names of sound designers in Europe, the main focus at this stage being the questionnaire elaboration and diffusion. The database of collected names results from several research paths among which personal relationships in the profession. In parallel, we have conducted an online research based on keywords, institutions, agencies, periodicals, websites and blogs (as shown in section 3.3). Finally, a ‘survey letter’ was sent to social networks (Facebook and LinkedIn), oriented mailing lists, personal contacts or other professionals (derived from conferences, associations, etc.) in order to enlarge the database, such as the ASD – Association of Sound Designers (sound designers in the UK theatre industry).

The resulting list (September 2018) comprises 708 names. However, several professionals cannot be considered as sound designers *per se*, as they rather qualify themselves, in many cases, audio engineers or acousticians. This decision depends upon choices made earlier while qualifying sound design (3.1) and defining a typical sound designer activity (3.2). All the skills/competences/facets that contribute to the conform definition need to be taken into consideration (although in different proportions).

After this ‘skimming’ process, ASDP’s data collection registered 558 professionals in sound design in Europe. This is obviously a work in process and we are still seeking professionals working in sub represented nations (bar chart 1). From the long-term perspective, our research will open to the entire world.

**4.2 The ASDP online questionnaire**

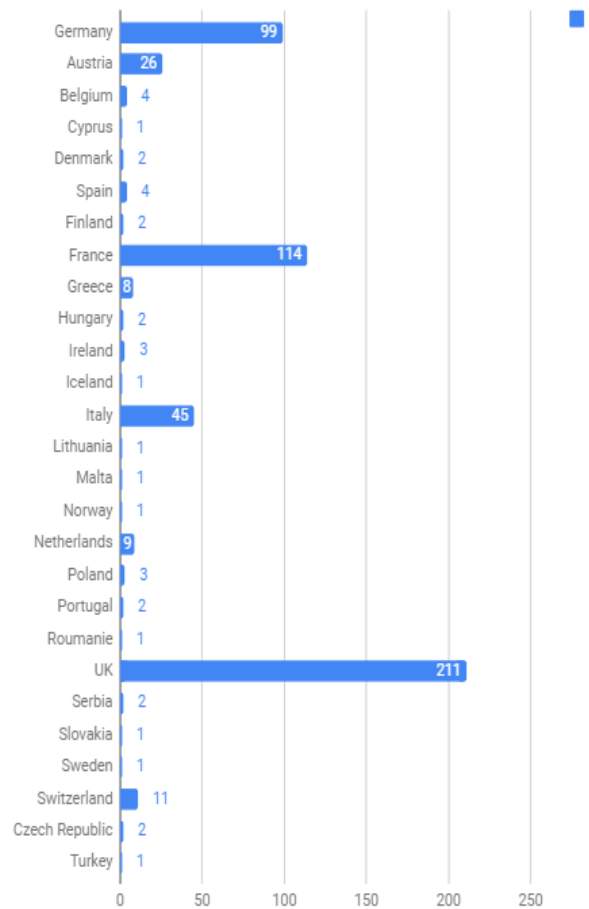
Just as a reminder, the primary purpose of this survey is to analyse and promote a better understanding of the role - identity and sound design practices - of sound designers in Europe.

An online questionnaire (section 5) was sent by email at the end of July 2018 - deadline to submission: September 30, to the above indicated list of professionals (section 4.1). Afterwards, a second reminder was sent during the second half of September.

**5. THE QUESTIONNAIRE**

The different questions (39 questions) result from a long process including coordination and role distribution of ASDP team members. We also based the

construction of this questionnaire on previous study or project that included interviews with professionals with regards to their methodology [43] and the use of specific techniques in sound production and prototyping [40].



**Bar Chart 1.** Inventory of sound designers in Europe (as of July 2018).

**5.1 Questionnaire structure**

The questionnaire is divided into 3 sections. The 1st section collects quantitative data. It consists of 19 questions, divided into two subsections. Part A concerns the participants’ personal profile (age, place of work, years and period of activity, background and training, areas and skills involved in each participants’ profession, work status). Part B analyses the participants’ teaching activity (if this is the case): environment, level, time dedicated to teaching, teaching methods. The second section, which is optional (17 questions, divided into two subsections), refers to ‘Methodology (Modus operandi)’ (Part C) and Technical Practice (part D). This section takes into account different criteria: working position / framework; role and recognition, timeframe and communication with stakeholders - duration of a project / phases; typical communication strategy, working methods - brainstorming, development, testing, revisions; personal archival of a project; hardware and

software environments; technical features in collecting sound. The third section (3 questions, optional) investigates general thoughts about sound design.

**5.2 Preliminary results**

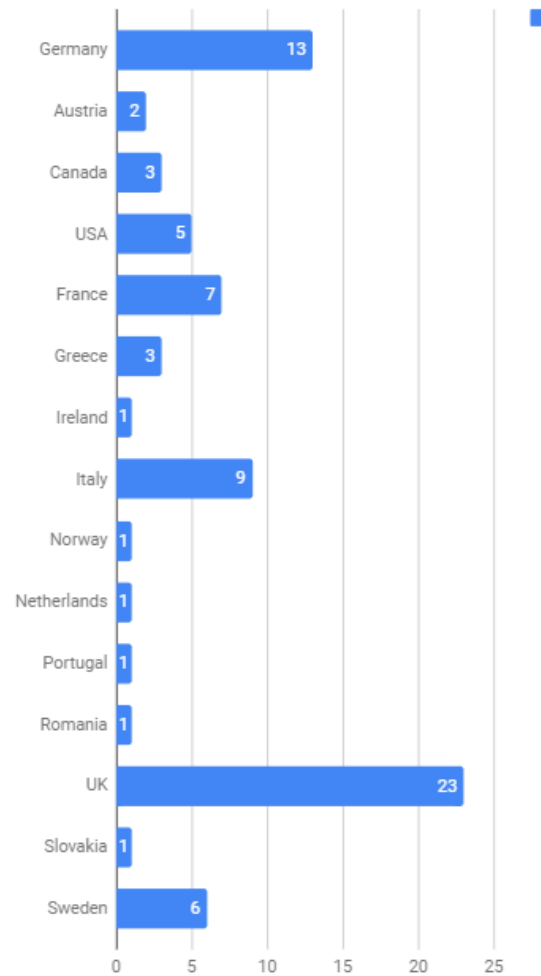
At the time this article was written (third week of September), we received 77 participations from the online questionnaire (bar chart 2). The majority comes from 3 representative countries: UK, Germany, Italy and France. This participation mirrors the number of professionals in our database. This is also due, one may think, to the presence of organized associations and recognized career paths in these nations.

So far we are able to analyse some quantitative data from the answers we've received. Bar chart 3 shows the distribution of age ranged from 18 to 25 years-old (seven respondents), 26-35 (seventeen), 36-45 (twenty-six), 46-55 (twenty), 56-65 (five). Such results already show that participants are adults well settled in their profession (36-55 years old), and also that there is a growing number of professionals among young people. Two participants are over 66 years old. If we compare these data with the increasing sound design literature over the last 15-20 years (section 2), we could easily assert that the sound design profession - working activity and related practices, has evolved and reached a new step in maturity.

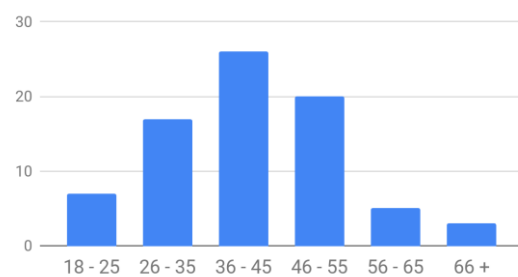
The survey data give information on the gender split (66 men vs. 9 women; 2 participants chose the 'I'd rather not answer this question' option) (pie chart 1). The sound designer profession still appears as a male profession, but there is a growing place for women (authors of [41] [42] are explicit figures in this domain).

Further analysis will involve text analysis (open questions), cross-analysis between different answers (within the same participant's thread of answers) and comparative analysis of data (and of course countries of origin). Question n.7 for example – What is your background and training? – offers insights with regard to the interdisciplinary background proper to the sound designer's profile mentioned earlier. Here is a list of skills learned during the training period (in decreasing order of occurrence): Audio engineering 19 (in which 3 are electronic engineers); Musical instrument training 14; Sound or electronic studio (electronic music courses or computer music courses, e.g. in universities or conservatories of music) 12; Music composition 11; Recording / Production 9; Film school and/or Video making 8; Self-taught in music performance (rock and popular music) 7; Sonic Interaction Design / Auditory Display 6; Arts & humanities 4; Musicology 3; Self-taught music/sound production (software) 3; Architecture and Urban Design 3; Drama School and theatre training 4; Physics 1; Psychology 1; Psychoacoustics 1; Law school 1; Biology 1; Social Communications and Advertising 1; DeeJaying 1. These preliminary data show that a combination of sound engineering and musical skills and training (instrument/composition) is common to sound designers. Such data also may be relevant to sound

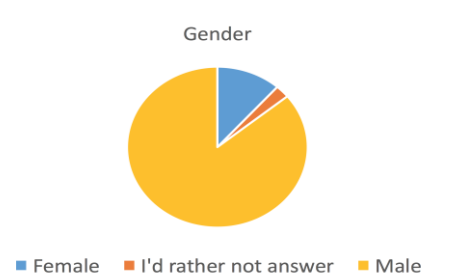
design program schools, pedagogical contents, etc



**Bar Chart 2.** Questionnaire participants (September 17 2018).



**Bar Chart 3.** Questionnaire participants (September 17 2018).



**Pie Chart 1.** Questionnaire participants (September 17 2018).

## 6. CONCLUSIONS AND PROJECTION

When this article was written (July-September 2018), the ASDP project was halfway through. The research results consist of numerical data - list of names, institutions, agencies, etc., from the European Sound Design community); an overview on “sound design” and “sound designers” translations in various languages, and preliminary results from the online questionnaire.

Semi-structured interviews will be recorded during the second half of the project October-February 2018. Face-to-face interviews will be made with professionals chosen with regards to their current work, their popularity or simply because they work in specific fields and relate to distinct practices. Academic research labs and relevant universities and/or schools are useful for epistemological purposes. We will also analyse sound design in cinema, theatre, museography, industry, sound branding and marketing.

We are planning to develop an original data search tool for dynamic cross-referenced information. This analytic procedure may allow a real time data representation based upon dynamic databases and 3D interfaces. By showing the results on a tridimensional scale (some preliminary tests has been developed during the last weeks) and by building up dynamic filters in real time, we plan to easily identify relevant connections between the overall data. Although the ASDP project is limited to the European community, the research of names and institutions we did develop is not exhaustive. It is our intention to promote our investigation in order to increase the professional network in the sound design industry. Moreover, future projects are planned for opening this research to the entire sound designers’ community.

One of our first ambitions was to reach as many persons as possible. Our mailing lists, social networks and personal acquaintances have partially reached this purpose (questionnaire answers received from outside Europe, as can be seen in bar chart 2, suggest the presence of an interest among the sound design community). However, we do acknowledge that most of the work towards a decent analysis of the sound designer’s practices require a tremendous effort in covering as much professionals as we can in Europe. In effect, another objective would be to attract the interest of the sound designers’ community and create a network for sharing scientific, artistic and professional knowledge, setting conferences and launching events. To this regard, we are presently creating a platform - website and blog: “thesounddesignproject.com” to collect more information - data, bibliographical references, and project realisation.

### REFERENCES

- [1] J.W. Humphrey, J.P. Oleson, A.N. Sherwood: *Greek and Roman technology: a sourcebook*, Routledge, 1998.
- [2] G. Dubus, R. Bresin: “A Systematic Review of Mapping Strategies for the Sonification of Physical Quantities”, *PLoS ONE*, 8 (12): e82491. doi:10.1371/journal.pone.0082491, 2013.
- [3] A. Farnetani, N. Prodi, R. Pompoli: “On the acoustics of ancient Greek and Roman theaters”, *The Journal of the Acoustical Society of America*, 124:3, pp. 1557-1567, 2008.
- [4] D. Collison: *The Sound of Theatre, A History - From the Ancient Greeks to the Modern Digital Age*, Plasa Limited, 2008.
- [5] C. Ebanista: “Paolino di Nola e l’introduzione della campana in Occidente”, *Dal fuoco all’aria. Tecniche, significati e prassi nell’uso delle campane dal Medioevo all’età Moderna* (Fabio Redi - Giovanna Petrella eds.), Pisa, pp. 325-353, 2007.
- [6] E. Quagliarini: *Costruzioni in legno nei teatri all’italiana del ’700 e ’800: il patrimonio nascosto dell’architettura teatrale marchigiana*, Alinea Editrice, 2008.
- [7] A. Farnell: *Designing Sound*, Cambridge, MA: MIT, 2010.
- [8] F. Napier, T. Guthrie: *Noises off: a handbook of sound effects*, Frederick Muller, London, 1936.
- [9] W. Whittington: *Sound Design and Science Fiction*, University of Texas Press, 2007.
- [10] R.M. Schafer: *The tuning of the world*, Alfred A. Knopf, 1977.
- [11] S. Vial: “Qu’est-ce que la recherche en design ? Introduction aux sciences du design”, *Sciences du Design*, (1), 2015, pp. 22-36.
- [12] M. Kilborn, “Some Advice for the Aspiring Sound Designer”, 18th 2013, [www.gamasutra.com/blogs/](http://www.gamasutra.com/blogs/).
- [13] L. Frankel, M. Racine: “The complex field of research: For design, through design, and about design”, *Proceedings of the Design Research Society (DRS) International Conference* (No. 043), 2010.
- [14] N. Cross: “Designerly ways of knowing: Design discipline versus design science”, *Design Issues*, 17(3), 2001, pp. 49-55.
- [15] M. Pryke, G. Rose, S. Whatmore: *Using Social Theory. Thinking Through Research*, Sage Publications – Open University, 2003.
- [16] B. Archer: “Design as a discipline”, *Design Studies*, 1(1), 1979, pp. 17-20.
- [17] N. Cross: *Designerly ways of knowing*, Springer Science & Business Media, 2007.
- [18] N. Misdariis, A. Cera, "Knowledge in Sound Design - The Silent Electric Vehicle —A Relevant Case Study", *Proceedings of the Conference on Design and Semantics of Form and Movement - Sense and Sensitivity*, DeSForM 2017, M. Bruns Alonso and E. Ozcan eds., pp. 185-195, 2017.
- [19] H. Wyatt, T. Amyes: *Audio Post Production for Television and Film - An Introduction to Technology and Techniques*, 3d ed. Oxford, Focal Press, 2005.

- [20] T. Gibbs: *The Fundamentals of Sonic Art and Sound Design*, Lausanne, Switzerland, AVA, 2007.
- [21] D. Kaye, J. LeBrecht: *Sound and Music for the Theatre*, 3d ed. Boston, Focal Press, 2009.
- [22] R. Brown: *Sound - A Reader in Theatre Practice*, Palgrave Macmillan, 2009.
- [23] D. Dal Palù, C. De Giorgi, B. Lerma, E. Buiatti: *Frontiers of Sound in Design. A Guide for the Development of Product Identity Through Sounds*, Springer International Publishing, 2018.
- [24] L. Zattra, N. Donin: "A questionnaire-based investigation of the skills and roles of Computer Music Designers, *Musicae Scientiae*, Special Issue Tracking the creative process in music, September 2016, vol. 20 no. 3, pp. 436-456.
- [25] F. Pecquet: "Nouvel ordre sonore, écoson et design musical: réflexions autour du projet de James Murphy pour le métro de New-York, [New sound order, ecosound and musical design: thoughts around the project of James Murphy for the New York subway]", forthcoming 2018.
- [26] E. Özcan, R. Van Egmond: "Product Sound Design: An InterDisciplinary Approach?", *Undisciplined! Design Research Society Conference 2008*, Sheffield Hallam University, Sheffield, UK (16-19 July 2008), 2009.
- [27] S. Vial: *Court traité du design*, 1ère éd. PUF (Paris), 2010.
- [28] R. C. Beckett: "Use of Wiki Tools in Collaborative Research and Learning", *Journal of Technologies in Knowledge Sharing*, 11 (4), pp. 13-26, 2016.
- [29] L. Vygotsky: *Mind in Society*, London, Harvard University Press, 1978.
- [30] M. M. Blattner, D. A. Sumikawa, R. M. Greenberg: "Earcons and Icons: Their Structure and Common Design Principles", *Human-Computer Interaction*, Volume 4, 1989, pp. 11-44, Lawrence Erlbaum Associates, Inc.
- [31] L. Kendrick: *Theatre Aurality*, Springer, 2017.
- [32] F. Pecquet: "Du design sonore (about sound design)", forthcoming, 2017.
- [33] G. D'Amico: "La professione del sound designer: MioJob.it intervista sounDesign", March 3, 2009, [www.soundesign.info](http://www.soundesign.info).
- [34] D. Sonnenschein (ed.): *Sound Design: The Expressive Power of Music, Voice, and Sound Effects in Cinema*, Michael Wiese Productions, 2001.
- [35] R. Sennett: *The Craftman*, Yale University Press, 2008.
- [36] P. Louridas: "Design as bricolage: anthropology meets design thinking", *Design Studies*, 20(6), 1999, pp. 517-535.
- [37] W. Rodriguez: *Le design sonore, naissance d'une catégorie musicale*, Mémoire de DEA, Ecole des Hautes Etudes en Sciences Sociales, EHESS, Paris. 2003.
- [38] D. Rocchesso, F. Fontata (eds.): *The sounding object*, Mondo Estremo, Firenze, 2003.
- [39] D. Rocchesso: *Explorations in Sonic Interaction Design*, Berlin, Logos Verlag, 2011.
- [40] S. Delle Monache: *SkAT-VG - Sketching Audio Technologies using Vocalizations and Gestures*, FP7-ICT-2013-C FET-Future Emerging Technologies-618067, Project Deliverable D7.7.2., <http://skatvg.iuav.it/>, 2017.
- [41] S. Pauletto (ed.): "Perspectives on Sound Design", *The new soundtrack*, Volume 4, Issue 2, September 2014.
- [42] K. Franinović, S. Serafin (eds.): *Sonic Interaction Design*, Cambridge, Massachusetts - London, England, The MIT Press, 2013.
- [43] D. Hug, N. Misdariis: "Towards a conceptual framework to integrate designerly and scientific sound design methods", *Proceedings of the 6th Audio Mostly Conference*, ACM, pp. 23-30, 2011.
- [44] P. Susini, O. Houix, N. Misdariis: "Sound design: an applied, experimental framework to study the perception of everyday sounds", *The New Soundtrack*, 4(2), pp. 103-121, 2014.

# Reflecting on the Role of Sound in an Immersive Multimedia Intervention for Health Communication

Sandra Pauletto

Department of Theatre, Film and TV, The University of  
York

sandra.pauletto@york.ac.uk

Bart Walus

Department of Theatre, Film and TV, The University of  
York

bartekwalus@hotmail.com

## ABSTRACT

We reflect on the role of sound in an immersive multimedia health communication intervention: *Jane's Story*. Health communication aims to inform, educate, and ultimately produce behavioral change (sometimes urgently, sometimes long-term). *Jane's Story* is an immersive eight minutes animation aimed at young people. It is produced as a tool to be used by educators to stimulate discussions around the topic of chronic diseases and disorders (diabetes, obesity, cerebral palsy, etc.) in adolescents. The intervention design merges visual methods typically used in health communication (infographics, data visualisation), with storytelling techniques typically found in filmmaking (first-person voiceover, use of music and sound design). We surveyed 37 subjects on the effectiveness of the intervention and the perceived role of its soundtrack. The analysis of the results reveals that the intervention was perceived as being generally effective, and that voice, music and sound design strongly support the communication of the health message and the emotional engagement of the audience with the topic. We then consider these results through the lens of Construal Level Theory, which describes how psychologically distant and urgent an issue is perceived to be depending on how it is communicated.

We then suggest that sound might be particularly effective in communicating concrete and urgent health messages that require a quick change, while symbolic visuals might be more effective portraying more general health messages, requiring long-term change. Additionally, this intervention shows that these two approaches can be combined effectively. Finally we report limitations and lessons learnt from this particular intervention that should be taken into account when designing multimedia health communication interventions.

Copyright: © 2018 Pauletto, S. et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution License 3.0 Unported](https://creativecommons.org/licenses/by/3.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

## 1. INTRODUCTION

Health communication can be defined as “the use of communication techniques and technologies to (positively) influence individuals, populations, and organisations for the purpose of promoting conditions conducive to human and environmental health.” [1] While the more immediate aims of health communication are to inform and educate, the final, more difficult, goal is to achieve behavioral change in the audience towards a healthier life. Communication methods have radically changed in the last twenty years. The most recent technological developments (sophisticated mobile technology, high definition audio and visual media, high quality sensors and interactivity) have opened up new possibilities. These advances have the potential to revolutionise the way in which health messages are communicated to both the individual and the general public.

The rapid expansion of fields such as eHealth [2] and persuasive technologies [3] confirm the momentousness of such area an of research, which in turn requires that we better understand the communication processes mediated by this technology and media.

Maibach and Holtgrave [1] have identified *entertainment education* (also known as educational entertainment or edutainment) as one of five distinct approaches to public health communication (others being social marketing, risk communication, behavioral decision theory, media advocacy, and interactive decision support systems).

Entertainment education is “the process of purposely designing and implementing a media message both to entertain and to educate, in order to increase audience members’ knowledge about an educational issue, create favorable attitudes, and change overt behaviours” [4]. Examples of media used in entertainment education are radio, television, popular music, theatrical performances, interactive computer software systems, and print materials such as comic books and magazines. Entertainment education is thought to have the poten-

tial to attract different audiences with the underlying assumption that people are more receptive to messages conveyed through entertaining experiences and that attention and reception might be higher in such contexts. Importantly, Singhal and Rogers [4] propose that the entertainment element can break down audience resistance to the educational content that, in itself, could be perceived as uninteresting and dull. It is therefore possible that health messages delivered in this way could more easily influence cognitive, affective, and behavioral outcomes. Specifically, positive outcomes could include a greater appreciation of the personal relevance of a problem, an increased knowledge and understanding of the problem, and an increased confidence in tackling the targeted issues. [5] A meta-analysis of published studies on the effects of entertainment education on health communication suggested that overall, entertainment education's effects on health outcomes – as measured by knowledge, attitudes, intention, and behaviors – was small but significant [6].

Examples of entertainment education can be found in family planning, substance abuse prevention, asthma, etc.

## 2. MULTIMEDIA IN HEALTH COMMUNICATION

A multimedia system delivers a message using a combination of different content forms such as audio, video, text, images, animations, etc. Many health entertainment education applications are multimedia (for example Public Service Announcements or TV programmes with a health message) and sometimes interactive (for example games with a health communication goal).

Two reviews of *Mass Media Interventions* (including television and radio) [7, 8] found that “mass media campaigns may have a positive influence upon the manner in which health services are utilized” despite “limitations relating to methodological quality and completeness of reporting of studies” [7, p.7]”, and that “their [mass media] true potential can be better understood, and pursued, by embracing an ecological framework of health that considers determinants of health across multiple fields of influence [8, p. 228]”, such as the individual, social network and community levels. Grilli *et al.* [7] also concluded that further work should focus, among other things, on “identifying characteristics of the message which influence the effectiveness of mass media campaigns” [7, p.7].

A review of *Interactive Health Communication Applications (IHCA) for people with chronic dis-*

*ease* [9] concluded that these applications “appeared to have largely positive effects on users, in that users tend to become more knowledgeable, feel better socially supported, and may have improved behavioural and clinical outcomes compared to non-users.” [9] The review also concluded that, although what makes such applications effective is still unclear, “it is likely that there are a number of factors contributing to IHCAs’ ability to improve knowledge, including the interactive nature of the information, the use of audio, graphics and video to present the information, and users’ ability to revisit the information repeatedly over time.” [9] Examples of successful interventions can be found in the area of childhood obesity [10], chronic back pain [11], to name but two. Despite the fact that extensive evaluation is lacking, overall these reviews conclude that, existing data on the effectiveness of multimedia (both interactive or not interactive) applications for health communication is promising.

### 2.1 Persuasiveness

Persuasiveness is a key characteristic of health communication as messages are designed with the deliberate intention to induce a voluntary change in attitudes and behaviours. Persuasion can be produced by the way a message is designed and presented. Research [12, 13] suggests that the presence of a structured narrative, both in interactive and non-interactive applications, can strongly support communication and persuasion. Although there is no general agreement on the definition of narrative, three main components are widely recognised as crucial in most narratives: the presence of a narrator’s voice, the existence of a conflict within a given context, and the description of actions taking place over time to resolve that conflict [14, 15]. The articulation of a message through a narrative structure “can captivate the audience, driving anticipation for plot resolution, thus becoming a self-motivating vehicle for information delivery. This quality gives narrative considerable power to explain complex phenomena and causal processes, and to create and reinforce memory traces for better recall and application over time.” [12, p. 13627] The medium used to convey the message, and its technology, can also contribute to the persuasiveness of a message. Fogg defines persuasive technology as “any interactive computing system designed to change people’s attitudes or behaviors.” [3, p. 1]. While Fogg places a lot of value on interactivity as a way to facilitate persuasiveness, we would argue that audiovisual immersion would produce a similar effect as it arguably

facilitates transportation and identification with the characters in the narrative. Overall, what seems to make a message persuasive is the degree of audience engagement created by the content/narrative and the way it is conveyed through the technology used.

### 3. THE DESIGN OF *JANE'S STORY*

*Jane's Story* was part of a public engagement project (Chronic health issues of adolescents: is the world listening?) funded by the Centre for Chronic Diseases and Disorders (C2D2, <https://www.york.ac.uk/c2d2/>) of the University of York and supported by the Wellcome Trust (<https://wellcome.ac.uk/>). The interdisciplinary team comprised health sciences researchers as well as sound and music theorists and practitioners and filmmakers. The authors produced the animation's soundtrack. The aim of the overall project was to develop an effective immersive multimedia entertainment education intervention that used video, sound (voice, music, sound effects and sonification) and immersion to communicate public health issues to young people, their parents, and educators. The piece focused specifically on communicating the social, mental and physical difficulties experienced by young people living with chronic diseases as well as disseminating global statistics about chronic conditions such as asthma, obesity, diabetes mellitus, and cerebral palsy. The data presented in the intervention were extracted from a World Health Organisation discussion paper *The Adolescent with a Chronic Condition* [16]. Specifically four data tables representing the prevalence of specific chronic diseases in young people in different countries in the world, are portrayed in the piece: Table 2: Prevalence of type 1 diabetes mellitus (p. 27); Table 3: Prevalence of asthma (p. 28); Table 4: Prevalence of cerebral palsy (p. 29); Table 5: Prevalence of overweight/obesity (p. 30). Additionally, excerpts from discussions in public internet fora where young people exchange experiences of living with chronic diseases, were quoted in the animation. The design of the intervention was influenced by two approaches: the use of infographics in health communication [17, 18, 19]; and the use of a fictional story, scripted on the basis of research data from interviews with young people, told from a first person point of view by *Jane*, the fictional narrator. The first-person narrative is a device used extensively in documentaries and television: it facilitates identification and empathy, and therefore engagement. The first person narrative approach was strongly recommended by the professional filmmakers in the project team.

The script describes the typical stages a young person goes through when they discover they have a chronic illness (some of the stages are: flare up, diagnosis, denial, anger, acceptance), but it does not identify any specific disease, signaling to the viewer that, although this is a story told in a first-person style, it is still fictional and general. In summary, we wanted to communicate general health information based on research in an engaging way and we did this by using the visual and the aural domains in quite different ways. The visual domain, with the use of infographics, text and diagrams, was used to remind the viewer of the general and objective nature of the information given; while the aural domain was used, to a large extent, to engage the viewer with the issues by facilitating identification and empathy. However some aspects of the soundtrack, the sonification of the data that accompanies the graphs presented towards the end, attempted to use sound in a more objective and general way. The animation (approximately 8 minutes long) was produced for a 3Sixty cube space with 4 four large screens and 32 speakers surrounding the audience.



Figure 1: People experiencing *Jane's Story* in the 3Sixty immersive space.

The immersive nature of the intervention was also designed to increase the engagement and focus on the health message. A single screen stereo version of the animation can be viewed at the following link:

<https://www.york.ac.uk/c2d2/projects/outreach/adolescents/>.

This version lacks immersion, but it gives a sense of the visual and aural elements used in the intervention. *Jane's Story* was exhibited a number of times both as an immersive audiovisual piece and in the cinema as a single screen piece. The audience varied between young people, health professionals and academics, teachers, and general public. Overall the reception has always been very positive.

A survey was carried out to attempt to identify how the soundtrack elements (voice, music, and sound design - sound effects and sonification) contribute to the communication of the health message

and the audience emotional engagement. The results are described below.

#### 4. SURVEY

A total of 37 subjects (12 male and 25 female) were surveyed (average age of 30.4, Standard Deviation (SD) 7.4). Thirty subjects declared having direct or indirect (through family or friends) experience of chronic disease and 7 subjects declared not having this experience. The subjects experienced the immersive intervention in groups of 4 or 5, and afterwards they were asked to fill in a questionnaire about the effectiveness of the intervention, and the role of the soundtrack in conveying the health message and emotionally engaging the audience. Subjects were aware that the animation was immersive, on the subject of chronic diseases, and were asked to not speak to each other at any moment during the piece and while answering the questionnaire. Subjects received a £8 voucher for their participation.

The results provide interesting insights for further study in this area.

##### 4.1 The overall animation

89% of the subjects found the film educational<sup>1</sup> and 70% stated that they would recommend the animation to others<sup>2</sup>. On a scale from 1 (not at all) - 5 (extremely effective) the animation effectiveness was rated 3.9 (SD 1.1).

Comments were generally positive and included “Excellent way to educate”; “To a younger audience - excellent; but could also benefit parents to help them understand”; “I think making such an emotive connection and seeing things from the girl’s perspective helped to convey the issues surrounding the health problem”.

A number of limitations were also highlighted: “I felt that there wasn’t really much by way of a conclusion”; “I wanted to know what particular illness I was learning about”; “Perhaps lack of clarity around diagnosis would stop audiences engaging”. The most recurrent feelings in the audience were: “curious - wanted to know what the condition was”; “sympathy”; “sadness”; “confusion”; “I experienced the fear and confusion of the character when she didn’t understand her condition”; “A desire to know what the condition was”.

The most recurrent aspect that the audience disliked about the intervention included: “I found it hard to know where to look with four “screens””;

<sup>1</sup> 8% didn’t find the film educational, and 3% didn’t know (one person did not answer).

<sup>2</sup> 22% were undecided, and 8% would not recommend the film (one person did not answer this question).

“It was not clear what the girl’s condition was. The message was not particularly clear at the end”; “I feel that I didn’t really understand the condition Jane suffered? Was that intentional to make it a more general story - or did I just miss it? It was confusing because I felt I was missing something, in a way to make it harder to understand.”

##### 4.2 The role of the soundtrack

We asked the subjects to score on a scale from 1 (none at all) – 5 (a great deal) how much each of the soundtrack components (voice, music, sound design) contributed to the delivery of the film overall message and to its emotional content.

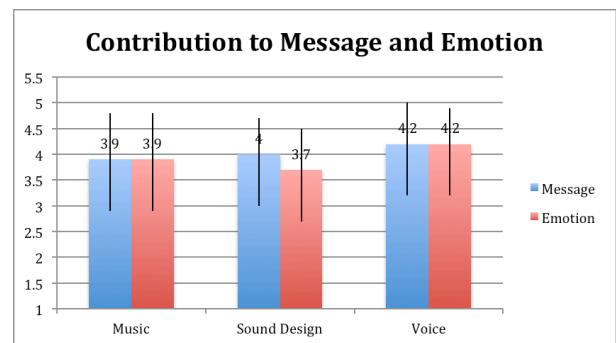


Figure 2: Music, sound design and voice contribution to health message’s communication and audience’s emotions

The Wilcoxon Signed-Rank Test indicates that, in terms of contributions to the message, the difference between voice and music results is significant ( $Z = -1.9777$ ,  $p = 0.0477 < 0.05$ ). The same test indicates that, in terms of contributions to the emotions, the difference between voice and music results is significant ( $Z = -2.1138$ ,  $p = 0.03486 < 0.05$ ) as well as the difference between voice and sound design results ( $Z = -2.9381$ ,  $p = 0.00328 < 0.05$ ).

The voice seems to contribute to the delivery of the message and the overall emotion the most, confirming the prominence of voice and speech in complex soundtracks. Then music and sound design seem to contribute slightly less, but very similarly, to the delivery of both message and emotion, with the sound design contributing slightly more in delivering the message than the emotions.

These results are confirmed by the most recurrent comments: “The music sets the mood of the video”; “It helped emphasize the mood throughout the story”; “And also to put the story into context, for example school/club”.

Most people attributed the most communicative impact to the voice, but attributed the portraying of emotions largely to the music: “The emotion was definitely carried forward by the music when you could feel the disappointment, anger and eventual acceptance and joy”. A few people perceived the

music as “not particularly emotive”; “un-invasive, but not boring”. They liked the music’s “simplicity” the fact that “it held my attention and felt contemporary”; “I neither liked or disliked the music really. Which is possibly a good thing as I remember the “message” more than the music”. Most people said about the sound design “They helped to visualise the situation when most of the animation was symbols”; “Contributed massively, enables you to lose yourself in the presentation”; “I think at times it may have detracted from the message. The beginning of the piece is deliberately ambiguous, and having such a bombardment of scary, intimidating noise shifts the focus away from listening to the narrator”.

Most people perceived that sound design had an emotive effect, but not as strongly as music or voice. “The ambient sounds helped convey emotion, but the effect of the music was much more significant.”; “I think sound effects reflect some stages in the story rather than influence the emotion of the listener”; “The sound effects carry myself or rather transport myself into her experience with her despair, turmoil and eventually her joy in breaking out from the depressive state that she was in”.

The spatialisation of the sound design and consequent immersive effect was liked by the audience as well as its narrative qualities: “It felt like it truly came from all around, not just several points”; “[I liked] the way the sound moved around you”; “It did have a narrative effect on its own. It could almost replace the voice narrative”; “It gave the presentation more depth”; “The sound design cleverly immersed the listener into the classroom and hospital scenes especially”.

Subjects confirmed the voice as the main carrier of the message and an important carrier of emotions: “The majority of the message was provided by the voice”; “The voice was very believable”; “I would not hear each word spoken by the narrator as the voice was spoken too quickly and softer than the music itself. Maybe because I’m not a native speaker but I was able to discern the message conveyed through the intonation.”; “The voice felt authentic, identifiable as Yorkshire/Northerner, making the presentation feel more intimate.”; “As if they were chatting with you”; “It made the words being said real - a real person”; “It gave the message a more personal feel and made me empathise more with the character”; “the voice was the main emotional feature”; “it makes the story feel real and personal, close to you”; “The voice was fairly similar throughout. Nearer the end the voice seemed to register more emotion as the character came to terms with her illness but the emphasis

seemed to be on the sound effects and music for me.”

The audience particularly liked the voice because it felt: “very believable”; “authentic”; “convincing”; for its “normalness”, [the] slight regional accent made me feel comfortable and gave me belief in the genuineness of the story”; “it felt genuine, sincere, believable and therefore was emotive”; “made me identify with her”.

### 4.3 The sonification

Towards the end of the animation, a number of graphs showing the prevalence of chronic illnesses around the world are shown graphically and also sonically through sonification. The graphs showed a globe turning with percentages of prevalence of particular diseases or disorders showing in each country, while the sonification audio output resembled that of a Geiger counter or a Morse signal. 60% of the subjects said they liked the sound; 35% were unsure whether they liked it or not; and 5 % did not like it. However only 32% of the subjects thought that there was a relationship between the data displayed on screen and the sonified data; 49% were not sure; and 19% did not think that there was a relation between sound and data.

These results are encouraging in regard to the type of sound chosen for the sonification, which was interpreted perhaps as reminiscent of data transmission between different places on earth. However, the sonification was not a clear display of the data. While in this intervention the clarity of the sonification as a display of information was not crucial, it is evident that, at least in this specific case, the visual display was more informative than the auditory display.

The comments about the sonification reflect this result: “I didn’t notice any particular sound design”; “I didn’t find the sound design for the last section especially memorable”; “I did not notice a specific relationship, but it worked well”; “I think it is timed in a similar fashion to when particular data points appear on screen”; “[it sounds] like lots of data crunching together. Matrix [the film] influence?”.

Sonification has been shown [24] to be potentially useful in health communication applications, especially in the context of mobile technology where visuals are restricted, however further research is needed to fully understand how to employ it effectively.

### 4.4 Interesting differences between subject groups

Twenty-four out of the 25 female subjects (96%) found the film educational, while only 8 out of the 12 male subjects (67%) found the film educational. 19 out of 25 female subjects (76%) would recom-

mend the film, while only 6 out of 12 (50%) male subjects would recommend it.

The score for overall effectiveness confirms this trend, with female subjects giving an average score of 4 (SD 0.9) and males of 3.5 (SD 1.2).

It is possible that the use of a female character in the film, Jane, facilitates identification and, consequently, engagement with the intervention in female audience members. Identification and therefore gender considerations, could have a considerable effect in health communication. This should be considered in the design of similar interventions.

In relation to this one subject commented: "I wonder how the gender of the voice/the main character influences the message perceived. Maybe it would be better to have a gender neutral voice or name." It is clearly quite impossible to have a "neutral" voice. Even in the rare occasions when voices are quite ambiguous, we tend to assign a gender to them in our minds. When a character is "materialized" through their voice it cannot easily be made general, un-gendered, and this has an effect on the audience identification and engagement.

Thirty of 37 subjects declared that they had experiences, either directly or indirectly, of chronic illness. Twenty-one of these 30 subjects (70%) said that they would recommend the film, while only 3 out of the 7 (43%) of the remaining subjects (with no experience of chronic diseases) would do the same. Additionally, the subjects with experience of chronic illness scored the effectiveness of the intervention 4 (SD 1), while the remaining subjects scored 3.4 (SD 1.2).

This seems to point to the fact that people with experience of a chronic disease, therefore with a personal involvement in the issues portrayed in the film, are more likely to think that it is important for others to watch a film of this kind. It could also suggest that people without experience of chronic illness are more difficult to convince and engage.

With regard to the contribution of the soundtrack elements (music, sound design voice) to the message and emotion, the subjects with no experience of chronic illness seemed to attribute more importance to music (average score 4.4 (SD 0.5)) than voice (average score 4 (SD 0.8)) to the elicitation of emotion. The same group attributed more importance to sound design (4.1 (SD 0.4)) than voice (3.7 (SD 1)) for the communication of the message. The group with experience of chronic illness follows the overall trend and attributes more importance to voice, rather than music and sound design, to the creation of message and emotion. These results could suggest that people without a strong personal involvement in the issues

represented pay more attention to relatively abstract elements such as music and sound design than voice. As 6 out of 7 (86%) subjects in this group found the film educational, if this could be found to be true, it could mean that, putting more attention in the creation of sound and music could be used to engage this group of people more, and increase their knowledge of the health issues described which are not familiar to them.

## 5. SOUND MAKES IT CLOSE AND PERSONAL

In this multimedia health communication intervention we tried to combine two different approaches, which reflected the expertise in the research team, one borrowed directly from health communication, using infographics and graphs to visualise health data, and one from television production, a first person-based storytelling approach which, in this case, is mainly realised through the soundtrack design and has the aim to engage the audience with the health message. The people who experienced the immersive piece seemed to generally like the approach taken and found it quite effective.

An analysis of the survey results with regard to the role of the soundtrack elements in this intervention reveals that sound is indeed, for the majority of the time, very effective in engaging the audience at an individual, personal and emotional level, but perhaps less so at portraying data. Sound is rarely emotion or gender neutral, we cannot easily escape it (we do not have ears lids) and, even when not explicitly spatialised, it is immersive because it "spreads" around the room. These characteristics, which were not fully at the forefront of our minds when designing the intervention, could perhaps be important if we look at them through the lens of Construal Level theory (CLT) [10, 21, 22], a psychological theory that provides a useful framework for designing effective health communication applications. One of the main aim of health communication is to reduce the perceived "distance" between the message and the audience. CLT proposes that the psychological distance to an object or event influences a person's thoughts and behavior in relation to that object or event. Psychological distance is assumed to have a number of interrelated dimensions (time, space, social distance, and hypotheticality) that together affect our mental construals of objects and events. CLT research show that people mentally construe objects that are psychologically near in terms of low-level, detailed, and contextualized features, whereas the same objects or events at a distance are construed in terms of high-level, abstract, de-contextualised

and stable characteristics. This way of thinking about events and objects affects how we make predictions, evaluations and affects our behaviour. Health messages often attempt to affect our behavior on at least two different levels: increase our perception of risk in situations that have negative outcomes (smoking increases the risk of cancer), and persuade us that generally positive and healthy behaviors are good for us (eating more fruit and vegetables is good for you). Chandran and Menon [23] have shown that a message presenting a negative outcome (a risk) as being close is more effective in inducing behavioral change than the same message presented as distant. For example, you need to stop smoking *now*, otherwise your risk of cancer will increase *tomorrow*. On the other hand, “persuasive arguments appealing to idealistic values appear to be more persuasive for temporally distant, as opposed to near, attitude objects.” [21 p. 12] For example, increase your intake of fruit and vegetables *in your life*, and you *will live longer*. Additionally, it has been shown that congruency between desired task and relevant message distance cues highly increases the effectiveness of the message [21]. For example, you need to stop smoking *now*, otherwise your risk of cancer will increase *in your life*, is not as effective as you need to stop smoking “*now*, otherwise your risk of cancer will increase *tomorrow*” in changing behavior quickly. Summarising, it seems that urgent behavioral changes are more easily elicited with a message that portrays the issue as psychologically close (temporally, socially, etc.), while long-term changes are more effectively presented with a message portraying issues as psychologically more distant. With this in mind, we suggest that sound, being immersive, often emotive and localised, is best at portraying the risk-related messages that need behavioral changes urgently; while visuals, particularly when they are used in their symbolic form, could be better at portraying messages that aim to induce long-term changes. This intervention uses both these approaches. The sound brings the problem close, it makes it concrete, and perhaps it would make a young person think about their immediate behavior towards a classmate with a chronic disease more than general information (graphs and text) alone; while the visuals provide a wider, more distant and general context that can perhaps influence the individual’s longer-term approach to the issues.

## 6. CONCLUSIONS

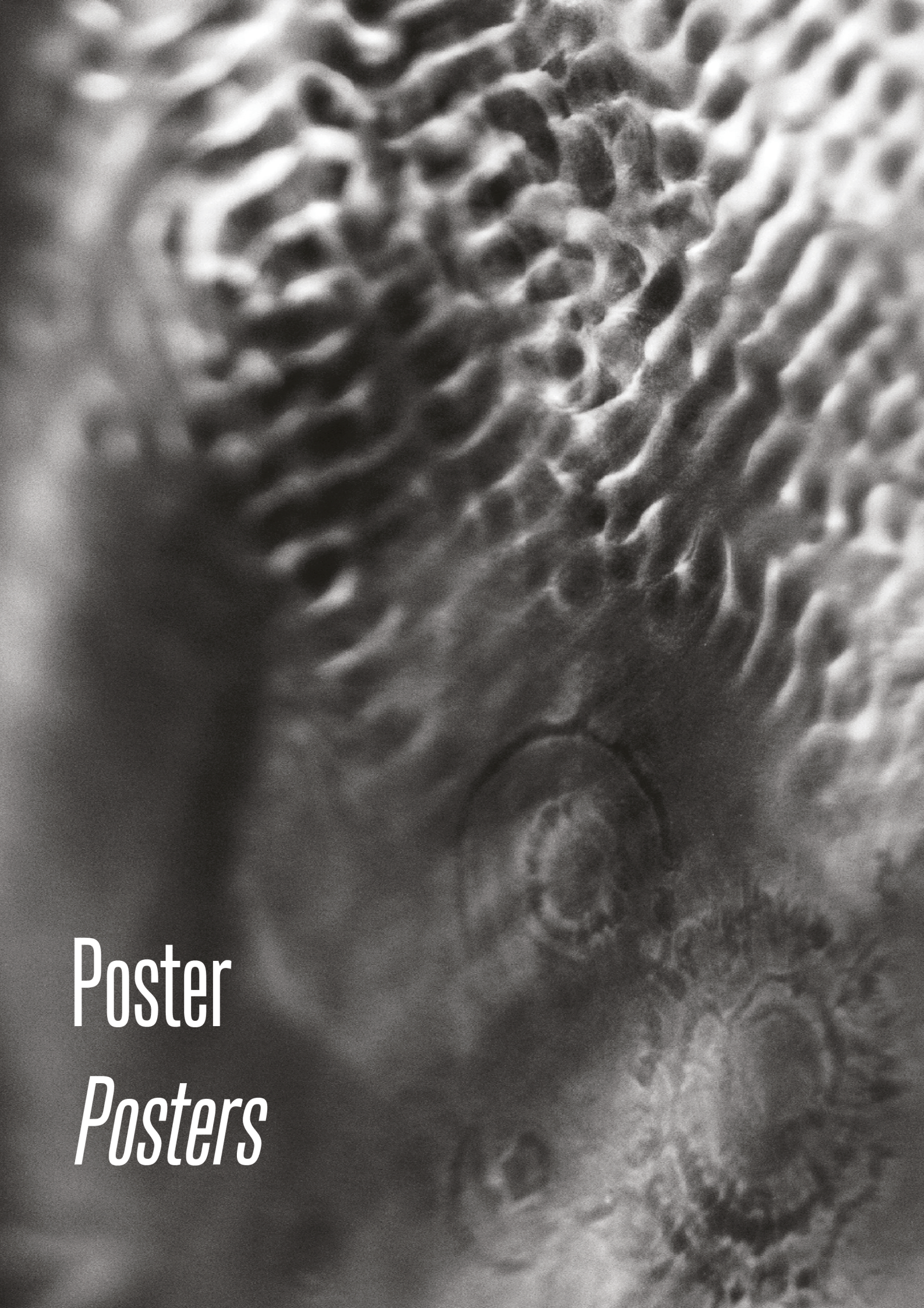
We have presented the design of a multimedia immersive health communication intervention and the results of a survey on its effectiveness and the role of the sound in it. Results indicate that sound has strong potential to make information concrete, close, and personal, facilitating identification, empathy and an emotional response. Reconsidering the design approach utilised in light of psychological theory that links the perceived distance of an issue with long-term or short-term behavioral changes, we suggest that sound might be particularly effective in communicating health messages that require urgent changes because it brings the message close and can make it more personal, while visuals, in particular symbolic visuals, might be better at communicating messages that require long-term changes. This intervention also suggests that the two can be combined, and produce an effective outcome. Further research would need to be carried out to fully investigate these suggestions.

## 7. REFERENCES

- [1] Maibach, E. and Holtgrave, D. R. Advances in Public Health Communication, *Annu. Rev. Public Health*. 1995. 16:219-38
- [2] Kreps, G. L. E-health: Technology mediated Health Communication, *Journal of Health Psychology*, 2003 8: 5, 5-6.
- [3] Fogg, B. J. *Persuasive Technology: Using Computers to Change What We Think and Do*, Morgan Kaufman Publishing, 2003.
- [4] Singhal, A. and Rogers, E. *Entertainment-Education: A Communication Strategy for Social Change*, Routledge, 1999
- [5] Bandura A. 1991. Social cognitive theory of mass communication. In *Media Effects: Advances in Theory and Research*, ed. J Bryant, D Zillman. Hillsdale, NJ: Lawrence Erlbaum
- [6] Shen, F. and Han, J., 2014. Effectiveness of entertainment education in communicating health information: A systematic review. *Asian Journal of Communication*, 24(6), pp.605-616.
- [7] Grilli, R., Ramsay, C. and Minozzi, S., 2002. Mass media interventions: effects on health services utilisation. *The Cochrane Library*.
- [8] Abrams, L. and Maibach, E. The Effectiveness of Mass Communication to Change Public Behavior, 2008, *Annu. Rev. Public Health*, 219-234
- [9] Murray E, Burns J, See Tai S, Lai R, Nazareth I, *Interactive Health Communication Applications*

- for people with chronic disease (Review), The Cochrane Library 2005, Issue 4.
- [10] Lu, A.S., Baranowski, J., Cullen, K.W., Jago, R., Thompson, D. and Baranowski, T., 2010. Interactive media for childhood obesity prevention. *Health communication*, 25(6-7), pp.581-582.
- [11] Frisch, A.L., Camerini, L. and Schulz, P.J., 2013. The impact of presentation style on the re-tention of online health information: a random-ized-controlled experiment. *Health communication*, 28(3), pp.286-293.
- [12] Downs, J.S., 2014. Prescriptive scientific narratives for communicating usable science. *Proceedings of the National Academy of Sciences*, 111(Supplement 4), pp.13627-13633.
- [13] Moyer - Gusé, E., 2008. Toward a theory of entertainment persuasion: Explaining the persuasive effects of entertainment - education messages. *Communication Theory*, 18(3), pp.407-425.
- [14] Hinyard, L.J. and Kreuter, M.W., 2007. Using narrative communication as a tool for health behavior change: a conceptual, theoretical, and empirical overview. *Health Education & Behavior*, 34(5), pp.777-792.
- [15] Bordwell, D., 2013. *Narration in the fiction film*. Routledge.
- [16] Michaud P-A., Suris J. C. and Viner R. 2007 *The Adolescent with a Chronic Condition*, WHO Discussion Papers on Adolescence, World Health Organisation, ISBN 978 92 4 159570 4
- [17] Zikmund-Fisher, B.J., Witteman, H.O., Dickson, M., Fuhrel-Forbis, A., Kahn, V.C., Exe, N.L., Valerio, M., Holtzman, L.G., Scherer, L.D. and Fagerlin, A., 2014. Blocks, ovals, or people? Icon type affects risk perceptions and recall of pictographs. *Medical decision making*, 34(4), pp.443-453.
- [18] Stones, C., 2017, November. *Developing Guidelines for Public Health Infographic Design*. In *European Journal of Public Health* (Vol. 27, pp. 7-7), Oxford University Press.
- [19] Catherine Stone <https://visualisinghealth.com>
- [20] Trope, Y. and Liberman, N., 2011. Construal level theory. *Handbook of theories of social psychology*, 1, pp.118-134.
- [21] Trope, Y., Liberman, N., & Wakslak, C. (2007). Construal levels and psychological distance: Effects on representation, prediction, evaluation, and behavior. *Journal of Consumer Psychology*, 17, 83–95. doi:10.1016/ S1057-7408(07)70013-X
- [22] Trope, Y., & Liberman, N. (2010). Construal-level theory of psychological distance. *Psychological Review*, 117, 440–463. doi:10.1037/a0018963
- [23] Chandran, S. and Menon, G., 2004. When a day means more than a year: Effects of temporal framing on judgments of health risk. *Journal of Consumer Research*, 31(2), pp.375-389.
- [24] Walus, B.P., Pauletto, S. and Mason-Jones, A., 2016. Sonification and music as support to the communication of alcohol-related health risks to young people. *Journal on Multimodal User Interfaces*, 10(3), pp.235-246.





Poster  
*Posters*



# bach: A Procedural Video Game Featuring Music Visualization for an Interdisciplinary Computer Based Approach to the Mathemusical Education

Giovanni Albini

Conservatorio "Jacopo Tomadini", Udine, Italy  
 Eesti Muusika- ja Teatriakadeemia, Tallinn, Estonia  
 mail@giovannialbini.it

## ABSTRACT

*bach*, a side-scroller video game developed by the author aiming to strengthen the interdisciplinary teaching of musical and mathematical knowledge, is introduced. Its gaming experience and the implications of its procedural level generation in the context of music visualization are then discussed in terms of their potential educational effectiveness with reference to experimentation results of a case study.

## 1. INTRODUCTION

Game-based learning environments provide a powerful resource for teaching and learning and have given rise in the last three decades to a specific and established area of research. It has been shown that "video games have the capacity to engage children in learning experiences", [1], and that they "not only teach content [...] but they may also teach us to learn, accelerating cognitive development", [2]. "Some evidence suggests that important skills may be built or reinforced by video games", [1], promoting "problem-solving, goal-oriented behavior, engagement and motivation", [3], and "research has shown how targeted games (i.e., games focused around particular concepts) might be used to create conceptual change", [4] But, although that should be all the more true in music learning, where the recreational element itself is foundational and underlined in many languages by musical words - *play*, *spielen*, *jouer*, etc. - it is well known that the recreational element is often overlooked in the teaching of music theory, [5], and that the potential of video games as tools to support and enrich formal music education have been mostly neglected, [6].

In this connection, the aim of this paper is to present *bach*, a video game designed by the author<sup>1</sup> to support interdisciplinary teaching of some fundamentals of counterpoint and Cartesian geometry. In fact, in [7] the author proposed an activity that was tested in 2007 for the second and third years of the junior secondary school in which

<sup>1</sup> *bach* has been developed in *Unity* game engine with *C#* language and built for Android, iOS, MacOS and Windows. It has been selected to be presented publicly for the first time at the third *Biennale d'Arte Contemporanea di Salerno*, Video Games and New Technologies of Art and Media section, October, 6<sup>th</sup> - November, 18<sup>th</sup> 2018.

Copyright: ©2018 Giovanni Albini et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution License 3.0 Unported](https://creativecommons.org/licenses/by/3.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

mathematics assisted music learning and conversely music theory supported learning mathematics, through a graphical visualization of Bach's works in the Cartesian plane. More specifically, it was targeted to "reinforce the [student's] ability to identify and analyze the perceptual properties of the musical sound through their abstraction and formal representation, [...] as well as to develop the mathematical skills in discerning invariant properties through the study of music and counterpoint"<sup>2</sup> [7]. *bach* has thus been designed with the goal to aid and strengthen such activity.

## 2. THE GAMEPLAY

The setting and the core of the gameplay of *bach* are briefly explained in its short tutorial: "*bach* [...] is a little bird who loves the music of J.S. Bach. In his dreams he flies over huge sounding cities born from the music of the German Maestro. Each note is a building: tall if the pitch is high and large if the note is long. He has to fly as close as possible to the skyline to truly enjoy the music he loves and to keep his dream afloat". MIDI files storing some of the keyboard music by Bach<sup>3</sup> are visualized as city skylines. *bach* moves horizontally at a constant speed (the tempo of the score) and while he flies over a building the associated note is played. Therefore, levels can be intended as a sort of city-like piano roll. Tapping and holding the screen makes *bach* ascend, otherwise he slowly loses altitude. A life-bar is given at the bottom of the screen: flying far from buildings, ascending and colliding consumes it, while flying close to the skyline replenishes it. If the life-bar runs out the player loses. The goal is to reach the end of the level (hence ending the listening of the score).

## 3. A CASE STUDY

To preliminarily test the effectiveness of *bach*, a study took place in two classes of the third year of a junior secondary school, 3A and 3B.<sup>4</sup> The activity described in [7] was condensed in two hours and carried out in the two classes independently. For pupils of class 3A the activity started introducing *bach* and always referring to it while introducing

<sup>2</sup> This author's translation.

<sup>3</sup> At present, they are: Inventions and Sinfonias, BWV 772-801, and Preludes and Fugues from the *The Well-Tempered Clavier*, BWV 846-881. However, the software can generate levels from any MIDI file.

<sup>4</sup> The activity took place on September, 21<sup>st</sup> 2018 in the classes 3A (25 pupils) and 3B (24 pupils) of the *Scuola Secondaria di Primo Grado "Pacifico Valussi"* in Udine. Professors Luigina Boschian, Doranna Di Vano and Mariaconcetta Squadrito assisted the author in carrying it out.

mathematical and musical concepts as well as their interrelations. For pupils of class 3B the content was the same but *bach* was not adopted. The pupils' response to the activity was then monitored via data collected from a final anonymous questionnaire. The first part of it was intended to assess their understanding of the concepts introduced through specific exercises and multiple-choice questions (with marks ranging from a minimum of zero to a maximum of twenty-three) while in the second part they were asked to state their satisfaction and engagement about the activity (totally ranging from zero to twelve). The average results for both classes are given in the following table that shows how class 3A, the one in which *bach* was introduced, got a much higher score in learning outcomes and a slightly higher one in satisfaction and engagement.

| Group | Learning outcomes | Satisfaction |
|-------|-------------------|--------------|
| 3A    | 17,80             | 10,04        |
| 3B    | 13,68             | 9,75         |

#### 4. DISCUSSION

The result of the aforesaid activity demonstrates how the audiovisual experience of *bach* could enhance the learning outcomes. This could be due to the following main reasons: because of the multimedia experience that brings together the music, its properties and their visual representations, thanks to the added value of the gaming experience, and due to the consequences of the player's goal-oriented behavior in the context of procedurally generated levels.

In fact, during the activity some pupils of class 3A asked the author and the teachers to play *bach* again and again, suggesting improvements for it with plenty of creativity. It seemed that the video game triggered their imagination: the experience that *bach* offered translated the abstract model of a geometrical representations of contrapuntal music into a dreamy and living setting. The music read on a score and heard and the geometries drawn in the Cartesian Plane truly came together and became one. Their link was directly experienced, felt, *played*, and hopefully remembered. Moreover, as underlined in [8], "regardless of the cognitive complexities evoked, games are essentially entertainment media and, as such, aesthetics play a large role in how they are realized and experienced. [...] What is aesthetically pleasing impacts our emotions and, in turn, our behavior".

In addition, in *bach* levels are generated from MIDI files with no human intervention.<sup>5</sup> One of the main features of such a procedural video game is the balance between unpredictability and expectations based on the consistent and meaningful generation of some aspects of the gameplay from the music. Therefore, in *bach* the player is motivated to reflect on how the music can affect its visual represen-

<sup>5</sup> Video games in which the gameplay is procedurally determined by a musical input are usually referred as music generated games. There are several examples of this genre: from the popular *Vib-Ribbon* (1999) for Sony PlayStation, that turned Audio CDs into levels of a rhythmical platformer, to the more recent niche experiment of *Adventures of General MIDI* (2014), an abandoned project of a platformer made from MIDI files built in Max/MSP.

tation, thus the levels and the gameplay. This could offer a further occasion to relate music properties to their geometrical counterparts for a better understanding of both. In fact, as noted in [9], "we rely on the practice of proceduralism to craft representations through rules, which in turn create possibility spaces that can be explored through play", and "to convey ideas effectively".

#### 5. A WORK IN PROGRESS

Although *bach* is fully functional, it is a constant work in progress. Several variations and prospective improvements of the gameplay, aiming at a more effective learning experience, are now being implemented and tested. For instance, the author is considering to introduce timed changes of the voice to follow<sup>6</sup>, eventually related to thematic content and contrapuntal relationships, other events connected to the musical content, such as for instance cadences, modulations and pedals, and color variations based on voices and tonalities.

#### 6. REFERENCES

- [1] M. Griffiths, "The educational benefits of videogames," in *Education and Health*, vol. 20 No.3, pp. 47-51, 2002.
- [2] H. J. Brown, *Videogames and Education: Humanistic Approaches to an Emergent Art Form*. M.E.Sharpe, 2008.
- [3] B. Gross, "Digital games in education: The design of game-based learning environments," in *Journal of Research on Technology in Education*, vol. 40, pp. 23-38, 2007.
- [4] K. Squire, "Open-ended video games: a model for developing learning for the interactive age," in *The ecology of games: Connecting youth, games and learning*, pp. 167-198, Cambridge: MIT Press, 2008.
- [5] C. Delfrati, *Fondamenti di Pedagogia Musicale*. EDT, 2008.
- [6] A. Paisley and G. Cassidy, "Music games: New opportunities for music education," in *Music, Technology, and Education: Critical Perspectives*, pp. 134-157, Oxford: Routledge, 2016.
- [7] G. Albini, "Musica e matematica: una proposta didattica," in *L'Insegnamento della Matematica e delle Scienze Integrate*, vol. 33 A No.1, pp. 65-83, 2010.
- [8] M. D. Dickey, *Aesthetics and Design for Game-based Learning*. Routledge, 2015.
- [9] I. Bogost, "The rhetoric of video games," in *The Ecology of Games: Connecting Youth, Games, and Learning*, pp. 117-140, The MIT Press, 2008.

<sup>6</sup> In fact, in the standard gameplay *bach* must stay close just to the highest notes of the generated skylines, regardless of the voice, so such an addition would help the player to focus on different voices in the listening of a contrapuntal score.

# PhOS: a photo-sensitive gestural surface

**Alberto Novello**  
 Conservatorio di Padova  
 albynovello@gmail.com

**Alessandro Fiordelmondo**  
 Conservatorio di Padova  
 dr.clipper@yahoo.it

**Arcangelo Di Donato**  
 Conservatorio di Padova  
 arcadido@gmail.com

## ABSTRACT

We present PhOS: a system that uses incident light intensity to control music parameters and sound synthesis. It is a transparent surface with 16 sensors each measuring ambient light, red, green and blue intensity values. The sensors are spaced on a 4x4 grid, making gestural control quite intuitive for the user, through the creation of light trajectories on the surface. Given its simplicity, with the direct translation of light into sound, and the ubiquitous distribution of light-emitting devices such as inexpensive bike lights and mobile phone flash lights, it can be used in multiple occasions from live performances to sound installations.

## 1. INTRODUCTION

For this research, we have been inspired by past contributions towards the interaction of light and sound, specifically when light is the agent responsible for the generation or modification of sound. We were particularly interested in the work of Austrian sound artist Peter Vogel and his cybernetic sound sculptures, which produce articulated and evolving rhythms from a few photocells positioned and soldered on a score-like grid [1]. We also took inspiration from the work of Klaus Filip, who rotates shutters of different shapes in front of fixed light beams to filter their signals and create a variety of luminous-driven sound patterns [2].

We found both these works especially successful because they don't require an explanation for the audience: through observation, one immediately imagines the relationship between light and sound and is driven to think possible creative developments of the system (e.g. by adding semi-transparent layers, or mirrors, or using alternative light sources).

However, we were interested in expanding the analog approach of the aforementioned authors, through the flexibility of digital technology and signal processing, by creating different sonic environments, or interaction paradigms, that respond in different ways to light stimulation. This would allow a performer to choose which music parameters to focus on: from lower to higher lev-

*Copyright: © 2018 First author's last name et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution License 3.0 Unported](https://creativecommons.org/licenses/by/3.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.*

el. We created PhOS a photo sensitive gestural surface made with a open source hardware and software.

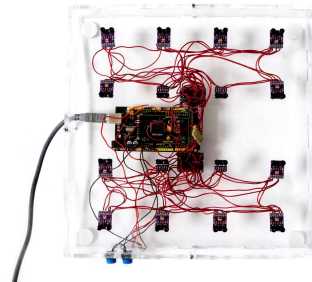


Figure 1. PhOS: photo-sensitive gestural surface

## 2. HARDWARE

We laser-cut a 30x30x4cm acrylic box containing 16 Arduino-compatible RGB sensors, positioned on a 4x4 matrix. The sensors use the I2C digital protocol of communication. Because a normal Arduino board can only receive one I2C channel, we used two I2C multiplexers, each managing eight I2C inputs, to receive the information from all 16 sensors. The Arduino Mega board inside the surface and its code receives the Ambient light intensity and the relative RGB levels from all 16 sensors. The surface sends all values with little latency to a computer via the USB cable, which is also responsible for the powering of the whole system (sensors, multiplexers and Arduino board). Six Pure Data patches loaded onto the computer produce sound from the light movements and variations. We added two buttons onto the surface side to navigate forward and backward through the patches, as well as a multicolor LED to show to the user the color relative to the active patch. The computer produces the sound through the available sound system (could be flexibly connected to stereo or multichannel system via cables or bluetooth). To maximize simplicity and playability, the patches were designed to be played without any computer screen, thus leaving all user attention for the light gestures. The computer can then be hidden away from the surface and becomes part of the sound system. In the context of installations, we also provide a set of inexpensive bike LED lights to create continuous or flickering light patterns, and a few black paper sheets to shutter parts of the surface from incident light. In this way, we encourage the participants to experiment with personal objects and explore furthermore the reactions of the surface to

new light conditions.

### 3. SOFTWARE

To avoid unwanted stimulation from background light, PhOS performs a minute-long self-calibration at startup, setting its noise gate level to the average light intensity of the surroundings. This method has proven to provide stable performance in the case of an open-air festival in which total darkness was not present. We also use a second calibration phase at start-up to set the thresholds for optimal light-flickering detection. The software does that iteratively finding the average value between dark and light intervals. All 4 data streams of each sensor (Ambient and RGB) are then low-passed with a 3ms window to remove the jitter deriving from internal noise or flickering artifacts that certain light sources produce. For each sensor we then analyze the presence of flickering vs continuous light using the thresholds. If flickering light is detected, we estimate its tempo: the time interval between consequent intensity onsets above the threshold. We then perform a grouping analysis, comparing values across different adjacent sensors. We define five regions: central, north-east, north-west, south-east, and south-west. If all four sensors of a region receive similar light patterns, the lights are grouped, the region-related switch is flipped and a new process or new decision can be triggered.

We then created the synthesis engine, consisting of different Pure Data patches. The main purpose for the creation of each patch was exploring different interaction scenarios to expand expressive possibilities, stimulate creative performative approaches and gestural dictionaries. We chose to use the different levels of RGB light for different synthesis strategies, the presence of flickering light to add accents in the sound, and trigger some cybernetic-inspired adaptive behaviors.

So far we've created five patches. The *RGB player* (blue Multicolor LED) reacts to the intensity of light playing sound samples from a disk. The light color individuates the sound bank, while the sensor individuates the specific sample. A reverb is added depending on how many sensors are excited simultaneously. The *Chaotic noise generator* (green Multicolor LED), inspired by Ciat-Lonbarde [3], uses the estimated coordinates of the mean light intensity on the surface to navigate through two chaotic oscillator frequencies and feedback amounts. It works best with one light at a time and reacts quite well to flickering light patterns creating gated noise rhythms. In the *Generative Techno Machine* (yellow Multicolor LED), each of the 16 sensors plays the harmonic partial of a drone with amplitude proportional to its incident light intensity. The RGB values enhance a particular aspect of sound, e.g. green controls noisiness, while red controls frequency modulation. The cybernetic behavior comes into play when a set of sensors detect flickering light: they might or might not decide to interact by adding percussive synthesized sounds. This decision depends on probabilities assigned to the patch which vary depending on what happened before. This keeps the patch voices in constant motion: even when a light is left flickering on the

surface, the resulting sound keeps on changing its harmonic and rhythmical structure, e.g., samples are substituted and harmonies rotated to create a dynamic sonic landscape, which can be immediately interrupted by moving away the lights from the surface. The fourth and fifth patch were conceived for an open air festival inside a natural reserve to explore the relationship of artificial and natural sounds (red and blue Multicolor LED). In both patches each sensor controls the sound level of a sample and the user can browse and interactively mix sound samples with light movements.

### 4. APPLICATIONS AND CONCLUSION

We used PhOS in two occasions so far: for performance and as installation in an open air festival. In both situations it caught the attention of the public for its simplicity and direct connection between light gesture and sound reaction. In the context of performance, PhOS is a practical interface to avoid the direct interaction with the computer screen, simplifying the connection of the performer to the audience. The use of light also makes it quicker for the performer to identify what's happening on the surface, thus allowing simultaneous handling of multiple sound events. Moreover the choreography of light movements constitute an extra layer to the performance. In the context of installation, the use of light control was quite intuitive for the audience, requiring no explanation of the instrument. We observed the audience enjoying playing together: sometimes children and adults would share the surface and "improvise" spontaneous sequences with it. Most users expressed enthusiastic comments even when the sound material was abstract and very bizarre for their everyday exposure.

Overall PhOS design expands previous light-controlled systems using multiple photocells distributed on a grid which allows for more complex interaction depending on gestural light trajectories, intuitive for users both in the context of performance and installation. The use of RGB levels also expands the expressive possibilities by simultaneously controlling a larger number of sound parameters. The inexpensive cost and the abundant availability of light sources make it easy to build and to share with the goal of seeing what different creative approaches users could extract from it. Due to the open software, it is also possible for the users to upload their own patches to personalize their instrument. Future improvements include the porting on Raspberry Pi [4], to have a self contained system that requires only an audio cable to output sound (or bluetooth) and a power supply.

A video example of PhOS can be watched here:

<https://youtu.be/nAIXPy88NHc>

### 5. REFERENCES

- [1] <http://www.petervogel-objekte.de/>
- [2] <http://filipino.klingt.org>
- [3] <http://www.ciat-lonbarde.net/>
- [4] <https://www.raspberrypi.org>

# IL SUONO ATTUALE DEI LUOGHI PERDUTI

**Arcangelo Di Donato**

Conservatorio “C. Pollini”, Padova  
arcadido@gmail.com

**Giorgio Klauer**

Conservatorio “B. Marcello”, Venezia  
giorgio@klauer.it

## ABSTRACT

Lo studio propone un esempio di dispositivo interattivo che permette l’esperienza differenziale di un soundscape attuale su una mappa antica. Concepito come un libro aumentato mediante interazione sonora e video, il dispositivo permette di scorrere il dito sulle aree sensibili di una cartografia della Padova rinascimentale, rivelando l’attuale paesaggio sonoro insieme a immagini. La differenza tra il soundscape attuale e la rappresentazione del paesaggio poco urbanizzato nel sedicesimo secolo determina una discrasia le cui valenze culturali sono poste in relazione all’ecologia acustica e gli aspetti qualitativi dell’esperienza al campo della progettazione urbanistica.

## 1. INTRODUZIONE

Padova era, fino a metà Ottocento, una città di canali navigabili che la attraversavano in ogni direzione, prevalentemente lungo le antiche mura del Castrum e delle mura cinquecentesche. Il tombinamento di queste arterie d’acqua, dalla metà dell’Ottocento fino al 1977 [1], quale intervento di “risanamento” e “modernizzazione” del centro storico, ha portato al completo stravolgimento del *genius loci*: una viabilità fluviale e pedonale ha fatto posto a strade di scorrimento per gli autoveicoli, con le conseguenze connesse all’inquinamento acustico e ambientale.

Dalle mappe cinquecentesche di Padova si evince un assetto urbanistico radiale che parte da un centro molto antropizzato fino ai margini agricoli sui bordi delle mura cinquecentesche [2]. Le emergenze architettoniche sono enfatizzate da ampi spazi liberi da costruzioni. Alla metà del ’700 le aree verdi all’interno delle mura patavine erano ben il 42% [3].

La situazione urbanistica attuale mostra una densità abitativa compatta, con minime discontinuità nel tessuto edilizio: oggi Padova è quinta tra le province italiane con più consumo del suolo [4].

## 2. METODO

Per evidenziare l’evoluzione dell’assetto del territorio l’urbanistica utilizza prevalentemente fonti indirette, più facilmente accessibili e complete, ricorrendo a quelle dirette

quando le prime risultano insufficienti per livello di approfondimento e specificità o per capacità di intercettare fenomeni e dinamiche territoriali insorgenti [5]. I principali bacini da cui attinge sono costituiti dalle banche dati degli istituti di statistica, dalle collezioni degli istituti cartografici, dalle biblioteche e dagli archivi di diverse istituzioni (statali, regionali, locali, ecclesiastiche). Il ventaglio delle fonti è diversificato per la loro natura ed è complessivamente ampio, comunque variabile rispetto al tipo di analisi e di strumento urbanistico [6].

Ad analoghe fonti accedono l’*historical soundscape* e il *literary soundscape*, recenti campi di ricerca interdisciplinare che hanno lo scopo di ricostruire il paesaggio sonoro in epoche passate. La ricostruzione può avvenire attraverso la documentazione delle attività musicali [7], riferendosi all’evoluzione delle qualità acustiche del patrimonio architettonico [8] o in base al materiale di archivio (audio, video, documentazione scritta, mappe ecc...) [9].

Nel presente lavoro si propone di applicare un metodo alternativo alla comparazione sinottica. La soluzione immaginata fa ricorso alla dissonanza cognitiva, per cui due manifestazioni contraddittorie vengono sovrapposte per far emergere le differenze. Quello della dissonanza cognitiva è un concetto introdotto da Leon Festinger per descrivere la condizione di individui le cui credenze, nozioni, opinioni contrastano tra loro (dissonanza per “incoerenza logica”) [10]. La dissonanza cognitiva nasce dal contrasto tra due atteggiamenti e produce una situazione di disagio che spinge l’individuo ad adottare tutte le possibili soluzioni per creare uno stato di coerenza, equilibrio e benessere.

La scelta del libro aumentato è stata determinata dalla sua presenza in scenari di utilizzo quotidiani e dai molteplici esempi proposti negli ultimi due decenni grazie all’evoluzione dei paradigmi del design dell’interazione che ha posto il libro al centro del discorso sull’interazione uomo-computer [11].

A partire da [12], attraverso i diversi prototipi la ricerca ha applicato le tecnologie nel libro aumentato nelle direzioni della realtà mista, aumentata e virtuale; in altri casi si è creato uno scenario di esperienza sui due livelli digitale-visivo e tattile-fisico; una terza tendenza, più eccentrica, si orienta a esperienze di adattamento e personalizzazione dove le caratteristiche percepite del libro non sono alterate significativamente.

L’efficacia comunicativa della sonificazione dei libri cartacei, in particolare nei libri per l’infanzia, è stata dimostrata in tempi recenti ed è legata al concetto del libro come *embodied interface* e al suono come fattore e risultato dell’interazione, ossia all’interazione guidata dal suo-

Copyright: ©2018 Arcangelo Di Donato et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution License 3.0 Unported](https://creativecommons.org/licenses/by/3.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

no attraverso controlli parametrici in tempo reale ricavati dall'interazione stessa [13].

Nel caso del dispositivo presentato in questo articolo si è preferito adottare una forma di interazione discreta per concentrarsi sulla dissonanza cognitiva causata dall'esperienza visiva e auditiva di oggetti già ben definiti e qualificati. Il fatto di confrontarsi con dei *field recording*, assieme all'importanza della componente mnemonica e della valenza ecologista dell'esperienza, hanno ulteriormente favorito questa scelta.

Preservare l'esperienza antica ed efficace di consultare una mappa (con la coordinazione visiva, manuale e tattile dello scorrere con il dito su un percorso) è stato alla base della scelta progettuale di mantenere quanto più possibile la relazione gestaltica tra i processi cognitivi e una percezione immediata del mondo fenomenico.

Attraverso la visione di una mappa del '500 combinata all'ascolto del soundscape attuale, si vuole restituire un'esperienza asincrona che induce il soggetto a desiderare un riequilibrio ecologico.

### 3. MAPPA DEI SUONI PERDUTI A PADOVA

Il progetto si basa sulla mappa di Padova inserita in "Civitates Orbis Terrarum" [14] pubblicato a Colonia nel 1572. Il fac-simile del volume alloggia, al di sotto della pagina che riporta la mappa, un circuito elettronico con sensori di forza collegati agli ingressi analogici di un microcontrollore connesso in seriale al PC (Figura 1). Il software gestisce l'interazione sul computer in modo che al tocco di un'area sensibile sulla mappa antica si visualizzi sullo schermo la posizione relativa alla cartografia odierna e si riproducano registrazioni sonore, immagini e filmati della situazione attuale.<sup>1</sup>

Il *field recording* è stato realizzato con un registratore portatile stereofonico in configurazione XY<sup>2</sup> a 170 cm di altezza e il medesimo guadagno in ciascun luogo. Le riprese fotografiche e video sono state effettuate a mano libera per mezzo di una fotocamera a ottiche intercambiabili.<sup>3</sup>

La vista della mappa del '500, combinata all'ascolto del soundscape attuale, costituisce l'esperienza principale, rispetto alla quale le informazioni iconiche e storiche presentate a video (mappa dei canali tombinati, foto e video associati ai punti di rilevamento acustico) sono accessorie. Una dimostrazione del dispositivo è visibile al sito <https://youtu.be/2QcvVYV-MkQ>.

#### 3.1 Luoghi

La scelta dei luoghi è stata dettata dalla necessità di individuare le emergenze di biofonia (suoni generati dagli esseri viventi diversi dall'uomo) sull'antropofonia (suoni generati dall'uomo, inclusi quelli elettromeccanici) per meglio



Figura 1. La mappa interattiva con il display visivo.

evidenziare, nel caso ci fossero, presenze di segnali sonori. Per i filmati, i rilievi fotografici e, in particolare, sonori sono stati scelti sei punti tra le mura antiche e quelle cinquecentesche, in corrispondenza dei canali antichi (Figura 2).

Il rilievo acustico, come quello architettonico, non è mai neutro, ma frutto di una scelta consapevole (il rilievo è un progetto). La scelta dei punti A, B ed E è data dalla loro collocazione in prossimità di canali rimasti integri dal '500. Nel punto D, benché su un canale integro, si è registrato in condizione di traffico normale per poterlo confrontare col punto C distante 100 metri.

#### A: Parco Fistomba

Nonostante la registrazione sia stata effettuata in un momento di scarso traffico, la vicinanza del ponte di Ognisanti, che collega il centro col nodo della Stanga, fa sì che dal rumore del traffico di fondo emergano solo debolmente i canti degli uccelli e le voci dei bambini che giocano nel parco.

#### B: Scalinata di Porta Portello

Questa scalinata, restaurata da alcuni anni, è un angolo intatto della memoria antica di Padova. Nella mappa cinquecentesca sono raffigurate barche all'ormeggio: qui c'è l'approdo del Burchiello che, oggi come in passato, collega Padova a Venezia. Nonostante il rumore di fondo del traffico stradale sia debole, dalla registrazione non risultano udibili segnali rilevanti.

#### C: Ingresso del Policlinico

In questo incrocio di canali tombinati regnano i suoni del traffico, saltuariamente dominati dalle sirene delle autoambulanze. Nel paesaggio sonoro non si ode alcun segnale.

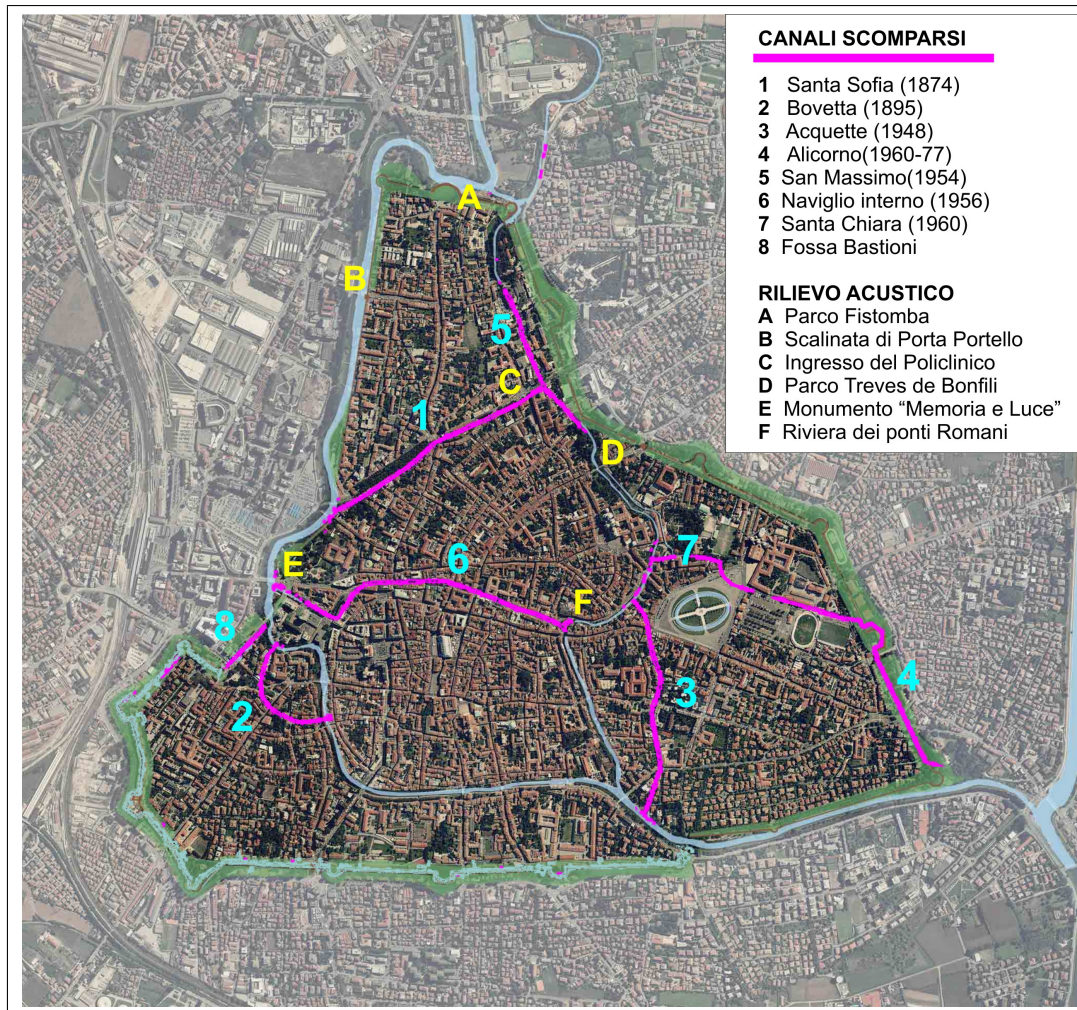
#### D: Parco Treves de Bonfili

Questo angolo di natura nel cuore di Padova fu progettato in stile romantico all'inglese dall'architetto Giuseppe Jappelli su commissione dei Treves de Bonfili e realizzato fra il 1829 ed il 1835. Il parco è molto piccolo, ma la sapiente disposizione di architetture fintamente antiche in un articolato percorso arricchito da piante esotiche, fa variare il paesaggio ad ogni passo. Qualifica la scenografia la presenza della via d'acqua che in questo tratto prende corsa

<sup>1</sup> Per interfacciare i sensori con il computer è stata utilizzata la piattaforma Arduino; come software di elaborazione del suono e per il *display* visivo sono stati usati Max e Jitter.

<sup>2</sup> <https://www.zoom-na.com/products/field-video-recording/field-recording/zoom-h5-handly-recorder#specs>

<sup>3</sup> <https://www.panasonic.com/it/consumer/fotocamere-e-videocamere/fotocamere-lumix-g.html>



**Figura 2.** La mappa di Padova: in giallo le antiche mura cinquecentesche, in azzurro i canali esistenti, in blu, numerati in rosso, i canali tombinati. Con le lettere sono segnati i punti di rilievo acustico. © Comitato Mura di Padova

facendo emergere il suono del ruscello. La combinazione del contesto arboreo, delle discontinuità del terreno e della presenza del canale permette di udire più distintamente il canto di uccelli, lo sciabordare sugli argini, i suoni di una voce e di chitarra emessi nel padiglione del parco, al di sopra del rumore del traffico e delle sirene del confinante policlinico.

*E: Monumento "Memoria e Luce"*

Questo monumento, progettato da Daniel Libeskind per ricordare i tragici eventi dell'11 settembre a New York, sorge in un luogo di incrocio di canali ancora esistenti. Sulla mappa cinquecentesca questo punto è una zona ai margini, poco o nulla antropizzata. La marginalità si respira ancora in questo angolo a ridosso della stazione ferroviaria. La registrazione effettuata in un momento di scarso traffico permette di udire minimamente segnali rilevanti, come il canto di uccelli.

*F: Riviera dei ponti Romani*

La dissonanza cognitiva regna nella stessa toponomastica: chi non conosce la storia di Padova si domanderà il senso della "riviera" nel punto di un incrocio di canali interrati. Rispetto ad altre situazioni simili, il rumore del traffico non impedisce di rilevare altri segnali perché ci si

trova in una zona a traffico limitato ai margini della zona pedonale. Caratterizza questo luogo il suono del tram su rotaia. Emergono voci, bici di passaggio e lo scampanio del tram elettrico.

#### 4. RIFERIMENTI CRITICI

##### 4.1 Analisi secondo il modello di Schafer

Le riprese sonore effettuate per il progetto sono state sottoposte preliminarmente a una analisi utilizzando le categorie proposte da Schafer [15].

*Toniche* (suono che evidenzia il carattere del luogo): il rumore dei motori degli automezzi e degli pneumatici sul manto d'asfalto sono predominanti in ciascuna registrazione di ogni singolo luogo.

*Segnali* (suoni in primo piano): elementi sonori caratterizzanti (canto di uccelli, sciabordio del ruscello, voci e suoni dei passanti) emergono solo in determinate circostanze (cfr. Sezione 3.1); la maggior parte del tempo tali elementi sono mascherati dal rumore sordo del traffico.<sup>4</sup>

<sup>4</sup> Vengono esclusi dagli elementi caratterizzanti il suono del passaggio ravvicinato di mezzi quali moto, tram, autobus e camion.

*Impronte sonore* (suono caratteristico di un'area): assenti.

## 4.2 Le esperienze sul campo di Krause

Di notevole interesse è, a riguardo di questa esperienza, la teoria di Bernie Krause, detta *Niche Hypothesis*, per cui i segnali acustici degli animali vengono modificati a seconda dell'ambiente in cui sono inseriti con l'obiettivo di massimizzare la loro propagazione all'interno dell'habitat [16, 17].

Un punto di incontro tra Schafer e Krause può essere la qualificazione di un paesaggio sonoro come Hi-Fi e Low-Fi, a seconda che abbia o meno un ricco spettro "armonico". I paesaggi Hi-Fi, in cui sono chiaramente distinguibili i diversi suoni e le loro caratteristiche (direzionalità, intensità, collocazione nello spazio acustico), consentono una percezione e una prospettiva acustica altamente definita; viceversa, i paesaggi Low-Fi sono scarsamente definiti, caratterizzati dal mascheramento di suoni sovrapposti, che comportano una prospettiva appiattita e una percezione confusa.

Dal punto di vista ecologico il paesaggio Hi-Fi è un paesaggio sano. Sotto il profilo antropico, per Schafer, segnali e impronte sonore ne arricchiscono l'esperienza; sotto il profilo biologico, per Krause, se il paesaggio è Hi-Fi vuol dire che ogni specie trova la sua giusta collocazione.

Sui caratteri qualificanti di un paesaggio Hi-Fi per individuare la ricchezza biofonica è illuminante l'esperienza di Krause citata in [18]: l'ascolto e l'osservazione dei sonogrammi effettuati a distanza di un anno nello stesso luogo, prima (Figura 3) e dopo (Figura 4) un intervento di deforestazione controllata, provano come l'habitat abbia subito un importante disagio nonostante l'intervento fosse volto a lasciare intatto il luogo alla vista (i sonogrammi rappresentano il drastico ridimensionamento del coro mattutino).

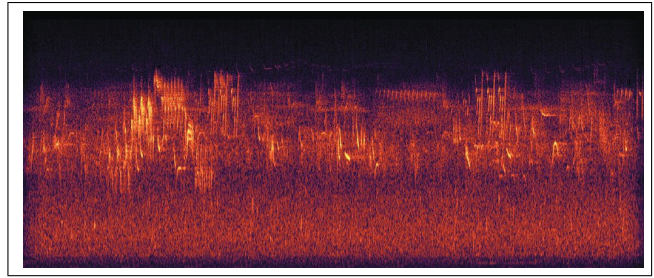
Il fenomeno può essere trasposto a Padova con la sostituzione delle vie d'acqua con strade e il conseguente decadimento da un paesaggio sonoro Hi-Fi ad uno Low-Fi.

Il primo genere di *soundscape* viene ritenuto tipico delle realtà preindustriali, mentre il secondo viene associato ai paesaggi prodotti dall'industrializzazione.

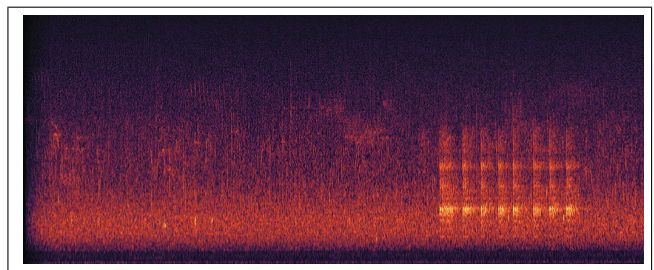
La vita cittadina, fin dall'antichità, è sempre stata rumorosa sia di notte che di giorno, come viene attestato, in un esempio di *literary soundscape*, da Marziale [19, libro IX epigramma 68] e Giovenale nella Roma antica (I sec. d.C). Già prima dell'alba gli schiavi iniziavano a far rumore pestando nel mortaio la farina per fare il pane, il gallo cantava dalla mattina fino all'ora sesta e forte si udiva il frastuono degli alunni della scuola elementare, che ripetevano in coro le massime enunciate dal maestro e urlavano quando venivano battuti o schiamazzavano durante le pause. Senza un attimo di sosta arrivava il suono delle martellate degli artigiani, il fracasso dei venditori ambulanti, dei cambiavalute, dei mendicanti e dei viaggiatori.

Questi rumori, per quanto fastidiosi, erano caratteristici solo delle grandi città ed in ogni caso possedevano una ricchezza timbrica e varietà degne di nota. Il moderno paesaggio sonoro Low-Fi è determinato dai rumori a bassa

frequenza originati da motori a scoppio e macchine elettromeccaniche. Subendo minore diffrazione, le basse frequenze tendono a permeare e saturare l'ambiente acustico, mascherando specialmente i fenomeni della biofonia.



**Figura 3.** Lincoln Meadow (California) prima dell'intervento di deforestazione, 1988. © Bernie Krause



**Figura 4.** Lincoln Meadow (California) un anno dopo la deforestazione selettiva, 1989. © Bernie Krause

## 5. SVILUPPI

Il far luce su un fenomeno quale pratica della progettazione urbanistica è denominato Valutazione Ambientale Strategica (VAS) [20]. Confluiscono nella VAS rilievi ambientali di varia natura; un processo circolare di analisi, predizione e implementazione porta alle scelte per la definizione del Piano di Assetto del Territorio (PAT).

In Italia, al momento, la mappatura acustica utilizzata per la VAS viene elaborata solo per il controllo delle soglie del rumore ammissibile. Tuttavia, la VAS è un processo che vede coinvolta la cittadinanza nel manifestare i bisogni percepiti come caratterizzanti la qualità della vita. Ciò considerato, sarebbe possibile e opportuno non solo monitorare l'inquinamento acustico ma anche far emergere gli elementi qualificanti del paesaggio sonoro introducendo metodologie ulteriori (ad esempio il *soundwalk*). Come testimoniato da recenti esperienze [21] l'indagine sulle qualità acustiche percepite nell'ambiente urbano (valutando registrazioni ambientali, *soundwalk* e verbalizzazioni) può costituire un riferimento per le decisioni di carattere urbanistico. Altri studi si spingono oltre proponendo come ausilio alla progettazione urbanistica metodi propri della composizione musicale [22].

Il presente contributo tuttavia è orientato a mettere in luce un fenomeno senza avere la pretesa di indagare sulle soluzioni, vista la complessità della materia urbanistica. Si fa notare come saltuariamente sorgano comitati spontanei che propongono lo stombinamento delle antiche arterie

d'acqua a Padova. Il presente lavoro può rientrare negli strumenti di buona pratica urbanistica per sensibilizzare la popolazione sull'evoluzione nel tempo delle caratteristiche acustiche della città e raccogliere proposte consapevoli.

## 6. CONCLUSIONI

“Mappa dei suoni perduti a Padova” si inserisce nel filone del *soundscape* ed inevitabilmente possiede una valenza ambientalista come gli studi di Schafer e Krause. Con questo lavoro si vuole introdurre un altro livello di lettura per mezzo di una mappa interattiva di cui si preserva l'estetica, lasciandola integra nella sua naturale collocazione di pagina all'interno di un libro in stampa anastatica. La qualità estetica del libro, in grande formato e stampa accurata, le aree sensibili al tocco del dito, rendono l'approccio alla lettura antico come la mappa rappresentata.<sup>5</sup> Scorrere il dito vicino ad un canale e sentire il rumore del traffico crea invece una discrasia che innesca un disagio e una riflessione su un paesaggio sonoro definitivamente perduto.

Al momento la mappa presenta solo sei punti di rilievo acustico ed è possibile immaginare una mappatura più estesa con una maggiore sensibilità al tocco. Il lavoro, al momento inedito, viene presentato al XXII CIM 2018.

## 7. BIBLIOGRAFIA

- [1] S. Lironi, “Edilizia popolare e urbanistica a padova nel novecento.” <https://www.slideshare.net/SergioLironi/sergio-lironi-edilizia-popolare-e-urbanistica-a-padova>, 2016. Accesso: luglio 2018.
- [2] S. Lironi and V. Ferrario, “Lo sviluppo urbano di padova, intervento seminariale nell'ambito del corso di diritto europeo dell'ambiente.” <https://www.slideshare.net/PaAMPadova/padova-sviluppo-urbano>, 2013. Accesso: luglio 2018.
- [3] E. Bevilacqua and L. Puppi, *Padova, il volto della città: dalla pianta del Valle al fotopiano*. Treviso: Editoriale Programma, 1987.
- [4] ISPRA, “Consumo di suolo, dinamiche territoriali e servizi ecosistemici.” <http://www.isprambiente.gov.it/files2017/area-stampa/dossier/consumo-di-suolo-2017/Consumodisuolosintesi2.pdf>, 2017. Accesso: luglio 2018.
- [5] L. Carbonara, *Le analisi urbanistiche. Riferimenti e metodi*. Roma: La Nuova Italia Scientifica, 1992.
- [6] C. Carozzi, G. Longhi, and R. Rozzi, *Popolazione, suolo, abitazioni. Introduzione all'analisi dei fenomeni urbani*. Padova: CEDAM, 1978.
- [7] “Paisajes sonoros históricos (c.1200-c.1800).” <http://historicalsoundscapes.com>. Accesso: luglio 2018.
- [8] E. Thompson, *The Soundscape of Modernity: Architectural Acoustics and the Culture of Listening in America, 1900-1933*. Cambridge (MS): MIT Press, 2002.
- [9] E. Thompson, “The roaring twenties.” <http://vectorsdev.usc.edu/NYCsound/777b.html>. Accesso: luglio 2018.
- [10] L. Festinger, *Teoria della dissonanza cognitiva*. Milano: Franco Angeli, 1997(9).
- [11] A. Pinto *et al.*, “Bridging books: the printed book as a support for digital experiences,” in *Proceedings of the 12th International Conference on Interaction Design and Children (IDC '13)*, pp. 606–609, 2013.
- [12] M. Back, J. Cohen, R. Gold, S. Harrison, and S. Minneman, “Listen reader: an electronically augmented paper-based book,” in *Proceedings of the SIGCHI conference on Human factors in computing systems (CHI '01)*, pp. 23–29, 2001.
- [13] S. Delle Monache, D. Rocchesso, J. Qi, L. Buechley, A. D. Götzen, and D. Cestaro, “Paper mechanisms for sonic interaction,” in *Proceedings of the Sixth International Conference on Tangible, Embedded and Embodied Interaction (TEI '12)*, pp. 61–68, 2012.
- [14] N. Grosso, *Le città d'Europa nel Rinascimento dal “Civitates orbis terrarum”*. Novara: De Agostini, 1995.
- [15] M. Schafer, *The Soundscape: Our Sonic Environment and the Tuning of the World*. New York: Simon and Schuster, 1977.
- [16] B. Krause, “The loss of natural soundscapes: ‘the divine music is growing dim’.” [http://www.earthisland.org/journal/index.php/eij/article/the\\_loss\\_of\\_natural\\_soundscapes](http://www.earthisland.org/journal/index.php/eij/article/the_loss_of_natural_soundscapes), 2012. Accesso: luglio 2018.
- [17] E. Stone, “Separating the noise from the noise: A finding in support of the ‘niche hypothesis’, that birds are influenced by human-induced noise in natural habitats,” *Anthrozoös*, vol. 13, no. 4, pp. 225–231, 2000.
- [18] “The great animal choir.” <https://beside.media/listening-to-the-great-animal-choir>, 2012. intervista a Bernie Krause e discussione con ascolto del “Bernie Krause records Lincoln Meadow after logging”. Accesso: luglio 2018.
- [19] F. W. Schneidewin, *M. Val. Martialis Epigrammaton libri*. Leipzig: Teubner, 1876.

<sup>5</sup> Di particolare interesse per le analogie con il presente progetto va citato “Aural Fabric”, che recupera il valore sensoriale di una mappa su stoffa per comunicare l'esperienza di un *soundwalk* [23].

- [20] C. Zoppi and M. Garano, *La valutazione ambientale strategica nella pianificazione territoriale*. Gangemi, 2003.
- [21] J. Kang, F. Aletta, E. Margaritis, and M. Yang, “A model for implementing soundscape maps in smart cities,” *Noise Mapping*, vol. 5, pp. 46–59, 2018.
- [22] S. Adhitya, *Sonifying Urban Rhythms: Towards the spatio-temporal composition of the urban environment*. PhD thesis, Università Iuav di Venezia / École des Hautes Études en Sciences Sociales, 2013.
- [23] A. Milo and J. D. Reiss, “Aural fabric,” in *Proceedings of the 12th International Audio Mostly Conference on Augmented and Participatory Sound and Music Experiences (AM '17)*, 2017.

# IL CONTROLLER PITTORICO: UNO STRUMENTO GESTUALE PER LA SINTESI SONORA

Arcangelo Di Donato

Conservatorio di Padova  
arcadido@gmail.com

Alberto Novello

Conservatorio di Padova  
albynovello@gmail.com

Alessandro Fiordelmondo

Conservatorio di Padova  
dr.clipper@yahoo.it

## ABSTRACT

Il “controller pittorico” permette di controllare la sintesi sonora dipingendo. Lo strumento consiste in una tavola il cui piano di lavoro è dotato di sensori che comunicano con un software musicale la scelta cromatica e la gestualità dell’artista tramite la conduttività elettrica dell’acqua e della grafite. Colori e segni vengono tradotti musicalmente in tempo reale.

## 1. INTRODUZIONE

Il rapporto tra pittura e musica è antico. L’approccio sinestetico è anteriore alla divisione tassonomica delle attività artistiche. Ma è in particolare con le avanguardie storiche del ‘900 che le barriere tra le arti vengono a cadere. Pittura e musica iniziano a condividere anche la stessa terminologia scambiandosi termini come composizione, colore, ritmo, gesto, ecc..

Tuttavia il presente progetto non va inteso come risposta materiale alle intuizioni espresse, in particolare, da Kandinskij e Schönberg nel loro epistolario [1] o negli scritti intorno al suono di Oskar Schlemmer ai tempi della docenza presso la Bauhaus [2]. In quelle esperienze si voleva mostrare come pittura e musica condividesse la stessa metodologia nell’ambito della composizione e anche come i risultati ritmici e cromatici fossero una *traduzione* dal linguaggio pittorico a quello musicale e viceversa.

Il presente lavoro fa riferimento più specificamente alle esperienze gestaltiche, cioè incentrate sulla percezione e sull’esperienza. Lo spazio percettivo, *gestaltico*, coinvolge fisiologicamente e olisticamente tutti i sensi, creando una fluidità delle informazioni tra i media. Un antesignano della fluidità delle percezioni sensoriali, in particolare tra pittura e suono, è stato John Cage.<sup>1</sup> Con l’avvento della tecnologia digitale si è approdati oggi all’idea della realtà aumentata, un unico spazio gestaltico dove confluiscono le diverse esperienze. Dalla composizione musicale e pittorica si è giunti alla composizione digitale<sup>2</sup>. Programmi come Max/MSP

<sup>1</sup> John Cage (1961) (a proposito dell’influenza di Robert Rauschenberg sulla sua estetica): “To Whom It May Concern: The white paintings came first; my silent piece came later.”[3].

<sup>2</sup> Lev Manovich sottolinea come nella composizione digitale si venga a creare una fluidità tra i media opposta all’estetica del

utilizzano gli stessi strumenti sia per l’audio che per il video. Questo permette una sorta di identità, se non altro progettuale, tra diverse discipline artistiche che oggi possono utilizzare la stessa strumentazione digitale.

## 2. LO STATO DELL’ARTE

Sono in vendita vernici e inchiostri (ad esempio *Circuit Scribe*) che permettono la creazione di circuiti su una superficie. Esistono inoltre alcune esperienze in cui si usa la grafite o colori conduttivi per realizzare opere che, al tocco di un elemento che chiuda il circuito, registri una resistenza variabile i cui valori sono associati a un suono. Tra le esperienze più vicine al controller proposto c’è il progetto *DrawAudio* di Jay Silver. Sul sito<sup>3</sup> dell’autore sono illustrate diverse soluzioni per creare musica con grafite e ogni elemento umido. Da questo progetto open source<sup>4</sup> è scaturito anche un prodotto commerciale destinato ai bambini (Fig 1).



Figura 1. *Brush with Genius*: prodotto commerciale del progetto open source *DrawAudio*

## 3. HARDWARE

Il controller pittorico si presenta come una tavola sul cui piano viene eseguita un’opera ad acquarello o a matita. L’acqua o la grafite fanno da conduttore. Sul

montaggio cinematografico: “Il montaggio mira a creare una dissonanza visiva, stilistica, semantica ed emotiva tra i vari elementi, invece la composizione digitale tende a miscelarli in un tutto integrato, un’unica gestalt” [4].

<sup>3</sup> <https://drawdio.com/>

<sup>4</sup> Il progetto completo è su: <https://learn.adafruit.com/drawdio/parts-list>

piano sono disposti dei sensori che registrano, tramite *Arduino Mega*, la variazione di resistenza quando il pennello o la grafite (collegati al polo positivo) chiudono il circuito. E' presente anche una tavolozza di acquerelli in pastiglie *godet* collegati anch'essi a dei sensori. La scelta del colore invia un segnale al software. Tutte le connessioni convergono in una scatola schermata che presenta dei led informativi per l'utilizzatore. Per i colori si è scelto di assegnare la sola funzione attivo/disattivo (come per un interruttore) mentre per il piano pittorico le resistenze sono variabili per poter mappare il percorso del pennello. Il pennello è un *Pentel Aquash Water Brush* ed ha un serbatoio pieno d'acqua. La punta, alla quale arriva un cavo collegato al polo positivo, è costantemente umida. Lo strumento è stato ottimizzato per il microcontrollore *Arduino* e al momento trova il suo software elettivo in *Max*. L'esecuzione può seguire un canovaccio (*Illustrare una favola*, [3.1]), improvvisare (*LLM 7 | Watercolour Scores*, [3.2]) o avere uno spartito di riferimento (*Violoncello ed elettronica*, [3.4]). L'allestimento per la performance artistica (Fig 2) prevede:

- Controller Pittorico, collegato ad un computer su cui giri un software;
- una webcam che riprenda l'attività pittorica;
- la proiezione dell'azione artistica;
- impianto di sonorizzazione.

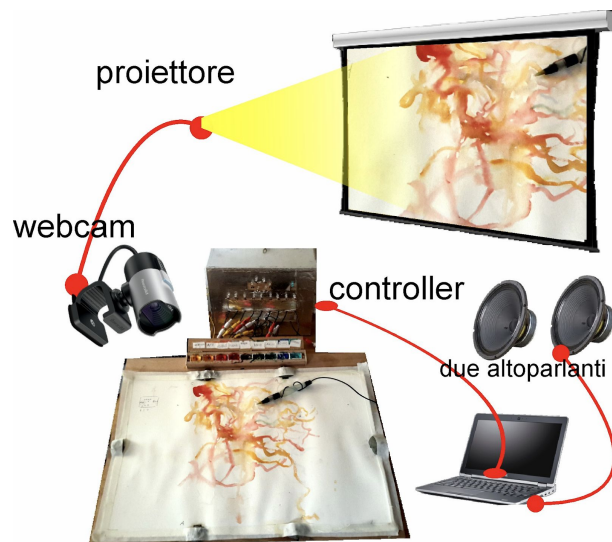


Figura 2. Schema dell'hardware per la performance con il controller pittorico.

## 4. APPLICAZIONI MUSICALI DEL CONTROLLER PITTORICO

### 4.1 Illustrare e sonorizzare la favola

Il controller pittorico nasce dal piacere di raccontare, illustrare e contemporaneamente sonorizzare la favola. Una prima applicazione in questo senso narrativo è

illustrata qui [https://youtu.be/nOR\\_amaXIKO](https://youtu.be/nOR_amaXIKO). Ogni colore seleziona una traccia audio (eventi atmosferici, suoni e rumori del bosco) e il gain è controllato dalla prossimità del pennello ai sensori.

### 4.2 LLM 7 | Watercolour Scores

La prima presentazione pubblica del controller pittorico è avvenuta nel giugno 2017 in occasione di un concerto per SaMPL<sup>5</sup> organizzato dal maestro Giorgio Klauer presso l'auditorium del conservatorio Pollini di Padova con le manipolazioni di Francesco Roberto Dani, Eric Guizzo, Matteo Polato e Nicola Sanguin. In questo caso l'azione pittorico/sonora entrava in feedback multimediale: la pittura generava suoni che venivano manipolati dai musicisti la cui musica influenzava a sua volta l'azione pittorica. E' possibile ascoltare e vederne il risultato qui: [vimeo.com/221398769](https://vimeo.com/221398769).

### 4.3 Controller audio/video

Il controller pittorico ha trovato una sua applicazione anche nell'ambito di una sonorizzazione video (progetto sviluppato nell'ambito di un corso tenuto dal M° Simone Faliva). In questa occasione si è creata una traccia in grafite (usando una matita 8B) che, percorsa da una matita (collegata al polo positivo), generava una resistenza variabile che controllava il flusso audio e video. Dal video pubblicato ([youtu.be/bSiJpvoyvOU](https://youtu.be/bSiJpvoyvOU)) si può osservare come il tracciato in grafite vada "seguito" come si segue uno spartito. L'esecutore-artista in questo caso esegue una partitura come qualunque strumentista.

### 4.4 Violoncello ed elettronica

Nel mese di Novembre 2017 si è tenuto il concorso "Violoncello ed elettronica" (coordinato dal M° Maurilio Cacciatore) presso il conservatorio di Castelfranco Veneto. In questa occasione l'attività pittorica controllava i parametri della granulazione applicata all'esecuzione del violoncellista Vittorio Ceccanti. Ad ogni colore era associato un preset diverso tramite *Max/MSP*.

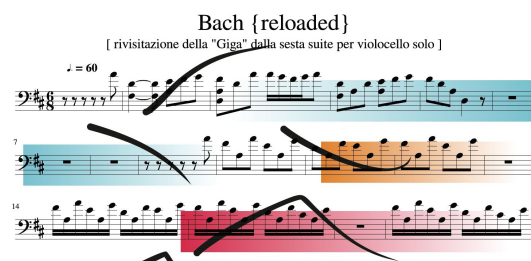
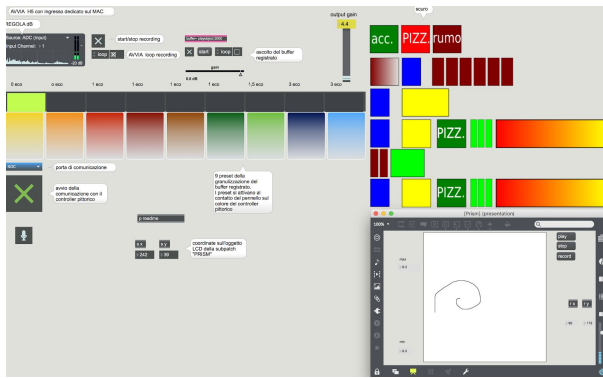


Figura 3. Dettaglio della partitura del controller per il concerto "Violoncello ed elettronica"

<sup>5</sup> SaMPL (Sound and Music Processing Lab) è un laboratorio integrato per la didattica, la ricerca e la produzione dedicate al Sound and Music Computing con sede a Padova.



**Figura 4.** La patch in “Presentation mode”. A sinistra: i preset associati alla *palette* dei colori usati. A destra: in alto lo spartito cromatico, in basso l’azione pittorica viene riportata su un “Display graphics” dove alle coordinate corrispondono valori di position randomness, grain duration, pitch randomness e pitch quantize.

#### 4.5 Azione pittorica e sonificazione

Il problema più rilevante nella proposizione della performance è quello di rendere trasparente il rapporto tra l’azione pittorica, la scelta cromatica e il risultato sonoro. C’è una vasta letteratura sul rapporto tra il colore e il timbro. Ci limitiamo in questa sede a rimandare a qualche riferimento bibliografico [5, 6] senza però sottrarci a delle considerazioni che condizionano le scelte pittorico-musicali dell’esecutore. La corrispondenza tra suono, forme, andamento lineare, note e colore è stato oggetto di considerazioni diverse nel corso dei secoli. Basta osservare la tabella in Fig. 5 per intuire come per gli artisti (in 3 secoli di storia di critica sinestetica) il rapporto tra note e colore fosse affidato alla personale sensibilità.

|                       |      | Three Centuries of Color Scales |        |        |       |      |        |      |       |       |      |            |             |
|-----------------------|------|---------------------------------|--------|--------|-------|------|--------|------|-------|-------|------|------------|-------------|
|                       |      | C                               | C#     | D      | D#    | E    | F      | F#   | G     | G#    | A    | A#         | B           |
| Isaac Newton          | 1704 | Red                             | Orange | Yellow | Green | Blue | Purple | Pink | White | Black | Grey | Light Blue | Light Green |
| Louis Bertrand Castel | 1734 | Red                             | Orange | Yellow | Green | Blue | Purple | Pink | White | Black | Grey | Light Blue | Light Green |
| George Field          | 1816 | Red                             | Orange | Yellow | Green | Blue | Purple | Pink | White | Black | Grey | Light Blue | Light Green |
| D. D. Jameson         | 1844 | Red                             | Orange | Yellow | Green | Blue | Purple | Pink | White | Black | Grey | Light Blue | Light Green |
| Theodor Seemann       | 1881 | Red                             | Orange | Yellow | Green | Blue | Purple | Pink | White | Black | Grey | Light Blue | Light Green |
| A. Wallace Rimington  | 1893 | Red                             | Orange | Yellow | Green | Blue | Purple | Pink | White | Black | Grey | Light Blue | Light Green |
| Bainbridge Bishop     | 1893 | Red                             | Orange | Yellow | Green | Blue | Purple | Pink | White | Black | Grey | Light Blue | Light Green |
| H. von Helmholtz      | 1910 | Red                             | Orange | Yellow | Green | Blue | Purple | Pink | White | Black | Grey | Light Blue | Light Green |
| Alexander Scriabin    | 1911 | Red                             | Orange | Yellow | Green | Blue | Purple | Pink | White | Black | Grey | Light Blue | Light Green |
| Adrian Bernard Klein  | 1930 | Red                             | Orange | Yellow | Green | Blue | Purple | Pink | White | Black | Grey | Light Blue | Light Green |
| August Aeppli         | 1940 | Red                             | Orange | Yellow | Green | Blue | Purple | Pink | White | Black | Grey | Light Blue | Light Green |
| I.J. Belmont          | 1944 | Red                             | Orange | Yellow | Green | Blue | Purple | Pink | White | Black | Grey | Light Blue | Light Green |
| Steve Zieverink       | 2004 | Red                             | Orange | Yellow | Green | Blue | Purple | Pink | White | Black | Grey | Light Blue | Light Green |

**Figura 5.** Tabella riassuntiva di varie corrispondenze tra note e colori (tatta da Fred Collop: *Color Scales?* [7]).

L’altezza è l’unico comune denominatore (di artisti<sup>6</sup>, musicisti e sentire comune) nell’attribuzione cromatica: suoni acuti=colori chiari e luminosi, suoni gravi=colori

<sup>6</sup> V. Kandinskij: “...il giallo diventa come una tromba acuta che squilli sempre più forte...” [12]

scuri e profondi. La ricerca di una mappatura con relazioni univoche tra colori e suono ha una ricca letteratura con approdi per lo più discordanti. Tuttavia oggi si può affermare con una certa sicurezza che in passato il problema della corrispondenza è stato mal posto. Le corrispondenze, se ci sono, devono essere ricercate tra attributi che si collocano su scale omogenee per tipo. In particolare J. L. Caivano [8] propone il confronto tra la saturazione (colore) e il timbro (suono) sulla base del modello HSL utilizzato per la definizione cromatica. Il sistema di coordinate dei modelli HSL (H =tonalità, S =saturazione, e L=luminosità) è basato sulla percezione che si ha di un colore in termini di tinta, sfumatura e tono. La saturazione è l’intensità di una tonalità specifica, che ci permette di discriminare tra una tonalità pura ed una tendente al grigio. I colori spettrali, derivati dalla scomposizione della luce attraverso un prisma, sono quelli che in assoluto ci appaiono più saturi, perché sono frutto - teoricamente - di un’unica lunghezza d’onda, mentre l’unione di diverse lunghezze d’onda produce un colore più smorto. Il timbro è una qualità del suono che dipende dal modo in cui le armoniche sono amalgamate ad un suono fondamentale. Da un punto di vista fisico quindi, Caivano collega il rumore alla luce bianca (che equivale all’insieme di tutte le lunghezze d’onda, rispettivamente sonore e luminose), il suono di uno strumento ad un colore di saturazione media (quindi l’insieme di alcune lunghezze d’onda) ed un’unica lunghezza d’onda ad un colore saturo. Inoltre, Caivano afferma che la singola frequenza viene percepita come “liscia”, così come la tonalità pura, mentre il rumore è percepito come “ruvido”, come per il colore grigio. Tuttavia lo studioso fa notare come in realtà la percezione del suono di uno strumento musicale risulti più “saturo” di una singola frequenza. L’analogia fisica, pur con questi limiti, traccia comunque delle linee guida importanti, da tenere in considerazione quando si parla di associazioni audiocolorate e ha portato Stephen Barrass ad orientare la ricerca nell’ambito dell’ *Auditory Information Design* [9]. Tuttavia l’applicazione di corrispondenze univoche colore/suono comporta una rigidità limitativa del processo creativo. L’azione pittorica del controller non vuole tradurre la scelta cromatica e formale in linguaggio musicale, ma fornire un aiuto uditivo nel comprendere l’azione pittorica. La scelta artistica di non seguire alcun criterio scientifico nella traduzione dell’opera pittorica in suono, abbinata ad una coerenza per cui la stessa scelta pittorica dà lo stesso risultato sonoro, è confortata dalla constatazione sperimentale che anche quando il suono non fornisce alcuna informazione utile migliora comunque l’accuratezza della visione. È infatti noto che il sistema nervoso combina le informazioni provenienti da diversi

riferimenti all'interno e attraverso le modalità sensoriali per migliorare le prestazioni nei compiti percettivi. In uno studio del 2011 condotto presso la *University of California* [10] si è rilevato che ai partecipanti sottoposti ad un compito di rilevamento di un movimento visivo, l'aggiunta di suoni che si muovevano nella stessa direzione ha migliorato l'accuratezza del giudizio. Recenti studi in ambito neurocognitivo hanno constatato che l'osservazione di azioni o comportamenti induce nel cervello dell'osservatore l'attivazione dei medesimi circuiti nervosi deputati a controllarne l'esecuzione, producendo una simulazione automatica definita *simulazione incarnata* [11]. Percepire un'azione non semplicemente come una sequenza di movimenti equivale a 'simularla' internamente, ossia ad attivare il suo programma motorio pur in assenza dell'esecuzione fattuale di quella stessa azione. Ciò consente all'osservatore di utilizzare le proprie risorse neurali per penetrare il mondo dell'altro 'dall'interno', mediante un meccanismo automatico di simulazione motoria. Grazie ai neuroni specchio si può spiegare l'empatia (*emfuhlung*) che si instaura tra artista e fruitore nel particolare momento della creazione.

Il corpo del fruitore, di fronte a opere che mettono in evidenza il gesto creatore dell'artista, reagisce quindi come se avesse compiuto lui stesso i movimenti necessari alla creazione. La simulazione incarnata simula mentalmente il processo che l'artista ha compiuto per creare l'opera.

## 5. CONCLUSIONI

### 5.1 Evoluzione dell'hardware

Le maggiori difficoltà costruttive riscontrate sono legate all'operare su di un hardware che basa il suo funzionamento sulla presenza dell'acqua (evapora, si addensa in alcuni punti, imbibisce troppo o troppo poco il foglio, corrode i circuiti ecc...). In futuro si potrebbe pensare di aumentare il numero dei sensori sul piano, per migliorare la mappatura del segno pittorico, e contestualmente migliorare l'aderenza del foglio sui sensori (che oggi è assicurata da alcune calamite). Inoltre sarebbe necessario avere tutte le terminazioni elettriche, che fanno contatto con l'acqua, dorate, perchè l'evento più frequente è l'ossidazione e la rottura dei cavi. Per enfatizzare il rapporto tra suoni e colori è possibile immaginare l'evoluzione dello strumento anche come controller di luci.

### 5.2 Evoluzione del software

Il suono ha un rapporto privilegiato con la dimensione temporale, ovvero descrive un fenomeno che si sviluppa nel tempo e la sonificazione è molto efficace per descrivere fenomeni che si espandono temporalmente: il controller pittorico vede l'opera d'arte come processo

e non solo come risultato. Il risultato dell'azione pittorica in questa performance musicale è uno spartito pittorico che documenta l'improvvisazione musicale. Con la stessa patch, ripercorrendo le stesse tracce e facendo le stesse scelte cromatiche si ottengono risultati analoghi sia pittoricamente che musicalmente. Il controller pittorico si è rivelato flessibile in diverse circostanze: sia nell'improvvisazione che seguendo delle partiture; sia in solo che in ensemble. Per il suo carattere di codifica in tempo reale il controller pittorico si avvicina al *live coding* lasciando in più traccia visibile di tutto lo svolgimento creativo. L'esecutore deve avere anche capacità pittoriche, ed è questo, oltre alla visualizzazione di un percorso sonoro, che suscita l'empatia degli (non più solo) uditori. Migliorare la corrispondenza tra azione pittorica e suono risulta essenziale per attuare la citata *simulazione incarnata* e migliorare la partecipazione empatica del pubblico.

## 6. REFERENCES

- [1] J. Hahl Koch, *Musica e pittura, Arnold Schönberg, Vasilij Kandinskij*, Ed Abscondita, 2018.
- [2] M. Mazzotta, *Kandinsky, Cage. Musica e spirituale nell'arte*. Ed. Skira 2017.
- [3] Brenda Lynne Leach, *Looking and Listening: Conversations between Modern Art and Music* Scarecrow Press, 2014, p.111.
- [4] L. Manovich, *Il Linguaggio dei nuovi media*, Milano, Olivares, 2001, p.187.
- [5] Marco De Biasi, *Breve storia del rapporto suono e colore*, 20 nov 2013, web <http://www.marcodebiasi.info/breve-storia-del-rapporto-suono-e-colore/>
- [6] Vasilij Kankinskij, *Scritti intorno alla musica*, a cura di Nilo Pucci, Discanto edizioni. 1979.
- [7] Fred Collop, *Color Scales?* 10/19/04, web. [rhythmiclight.com/archives/ideas/colorscapes.html](http://rhythmiclight.com/archives/ideas/colorscapes.html)
- [8] José Luis Caivano, *Color and sound: physical and psychophysical relations*, rivista: Color Research & Application, Volume 19, Numero 2, 1994
- [9] Stephen Barrass, *Auditory information design*, Doctoral Dissertation, The Australian National University (Australia), 1997, ISBN:0-599-37487-X
- [10] Kim R1, Peters MA, Shams L., *0 + 1 > 1: How adding noninformative sound improves performance on a visual task*, Research Article, 2011, <https://doi.org/10.1177/0956797611420662>
- [11] [http://www.treccani.it/enciclopedia/simulazione-incarnata\\_%28Lessico-del-XXI-Secolo%29/](http://www.treccani.it/enciclopedia/simulazione-incarnata_%28Lessico-del-XXI-Secolo%29/)

- [12] Vasilij Kandinskij, *Lo spirituale nell'arte*  
ed. SE, 2005, p. 29.

# SAFE AND SOUND

**Alessandro Barile**

Conservatorio G. Tartini Trieste  
barilealessandro3@gmail.com

**Samuel Montalto**

Conservatorio G. Tartini Trieste  
montalto.samuel@gmail.com

**Luca Cigaina**

Conservatorio G. Tartini Trieste  
lucacigaina@gmail.com

## ABSTRACT

Con “Safe and Sound”, versione inglese del nostro modo di dire “sano e salvo”, ci riferiamo ad un sistema che mira a rendere più sicuri gli elettrodomestici di nuova generazione fornendo un monitoraggio sonoro dello stato operativo dell'elettrodomestico e della sua interazione col fruitore.

La tecnologia ha puntato alla produzione di strumenti sempre più silenziosi nei loro meccanismi di funzionamento e a causa di questo processo sono state eliminate le informazioni legate al suono acustico, naturale, meccanico dello strumento stesso. La tecnologia si può comunque usare per implementare un sistema di monitoraggio audio del comportamento dello strumento capace di restituire informazioni di utilizzo.

In questo lavoro abbiamo applicato il concetto di “Safe and Sound” ad un piano di cottura ad induzione e ad un ferro da stiro.

## 1. INTRODUZIONE

Il monitoraggio di un'azione o di un evento attraverso il suono è un processo a cui qualsiasi persona è abituata, sia per quanto riguarda contesti naturali che artificiali.

Il malfunzionamento di uno strumento meccanico è facilmente riscontrabile se esso produce un suono diverso dal solito, per non parlare dei suoni chiaramente associati al pericolo, come ad esempio il rombo di un motore che percepiamo avvicinarsi rapidamente. Queste considerazioni fanno capo all'ambito dell'*ecological approach* di Bill Gaver [1], in particolare ci si riferisce al fenomeno definito *everyday listening* dove l'orecchio, più che al suono in sé, pone attenzione all'evento che gli è legato.

Sono gli stessi strumenti che noi costruiamo, con i loro suoni legati ai loro meccanismi, a darci informazioni su cosa stia accadendo e sul come. L'assenza di suono corrisponde ad una mancata informazione ed anche un' eccessiva attenuazione può portare ad un'interpretazione errata.

Questo tipo di fenomeni potrebbe influenzare negativamente l'utilizzo dello strumento, sia in termini di effettiva efficacia dell'operazione, sia di percezione ed esperienza di utilizzo.

*Copyright: © 2018 Barile et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution License 3.0 Unported](https://creativecommons.org/licenses/by/3.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.*

Si pensi alla riproduzione di campioni audio per indicare lo scatto di una macchina fotografica digitale. Il suono dà un'immediata conferma del successo dell'operazione, la quale diviene in qualche modo più “reale”. Similmente un aspirapolvere troppo silenzioso può essere percepito come poco efficiente.

La ricerca svolta nell'ambito del *Sonic Interaction Design (SID)* si occupa anche di questo aspetto, ossia quello di fornire, per così dire, una voce a strumenti “tecnologicamente muti”. Il nostro progetto è stato sviluppato all'interno del corso SID del Conservatorio G. Tartini di Trieste tenuto dal Prof. Pietro Polotti.

## 2. SORGENTE D'INFORMAZIONE

I suoni di monitoraggio, ossia specifici suoni fatti corrispondere a specifici eventi, vengono prodotti in concomitanza ed in accordo con le gestualità ed i processi di funzionamento dello strumento.

Nel caso del ferro da stiro sono stati applicati due sensori di temperatura, uno per la temperatura dell'acqua ed uno per la temperatura del ferro. Queste informazioni vengono tradotte in segnale audio per segnalare all'utente il livello termico dello strumento nelle sue parti ed il raggiungimento della condizione ottimale di utilizzo. Sull' impugnatura ergonomica del ferro da stiro è stato applicato un sensore di flessione, in modo da poter distinguere mediante monitoraggio audio le fasi di utilizzo dell'elettrodomestico, ovvero di attenzione e controllo sull'azione o meno. In altre parole si monitora il momento in cui l'utente cessa di utilizzare la macchina, o addirittura se ne allontana mentre questa è ancora funzionante. In aggiunta un giroscopio permette di distinguere fra la fase orizzontale e quella verticale. Ad esempio grazie a questa combinazione di elementi è possibile segnalare acusticamente che il ferro da stiro ancora caldo è stato lasciato per troppo tempo inutilizzato in posizione orizzontale.

Similmente, nella piastra a induzione un sensore di temperatura trasmette l'informazione che viene poi tradotta in segnale di monitoraggio.

## 3. SENSORISTICA E SONIFICAZIONE

### 3.1 Ambienti

Per la realizzazione di questo progetto sono stati utilizzati l' ambiente arduino per rilevare le informazioni

trasmesse dai vari sensori e Max/Msp/Live per la sonificazione.

### 3.2 Ferro da stiro

Al ferro da stiro è stato applicato un sensore di temperatura in prossimità del ferro, in modo tale da non danneggiarlo una volta raggiunte temperature troppo elevate, come si vede in figura 1. I valori di temperatura forniti dal sensore sono quindi stati calibrati in relazione allo stato del ferro, trovando così la soglia che corrisponde al raggiungimento della temperatura di utilizzo.

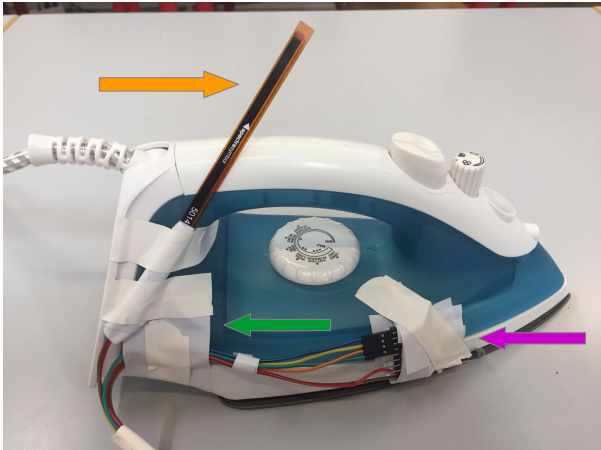


Figura 1. Il ferro da stiro, in fase di testing, con i sensori di flessione, temperatura e giroscopio.

Una volta oltrepassata tale soglia, viene prodotto un suono di sintesi che segnala lo stato operativo del ferro. Questo segnale viene costantemente emesso finché la temperatura non scende sotto alla soglia stabilita o fino a quando il ferro da stiro non viene impugnato, azione resa riconoscibile dalla variazione del valore rilevato dal sensore di flessione applicato al manico.

Il giroscopio applicato al ferro opera una riduzione dell'ampiezza del segnale audio di monitoring quando lo strumento è posto in posizione verticale, ossia in posizione di pausa.

Se il ferro da stiro viene lasciato in posizione orizzontale per più di 5 secondi senza essere impugnato, allora viene generato un segnale d'allarme per segnalare il pericolo di bruciare il capo d'abbigliamento.

Anche il riempimento della caldaia e la temperatura dell'acqua sono monitorabili. Il primo tramite un sensore di prossimità che rilevi l'innalzarsi del livello dell'acqua, il secondo tramite un sensore di temperatura.

Questo aspetto è stato simulato mediante un sensore esterno, non essendo stato possibile allo stato attuale integrare il sensore nel serbatoio d'acqua del ferro da stiro.

### 3.3 Piano di cottura

Al piano di cottura è stato aggiunto un bottone d'azionamento, come si vede in figura 2, a cui è stato associato un suono specifico per indicare la corretta accensione della piastra. In seguito all'accensione viene prodotto un

altro suono di sintesi che serve per segnalare lo stato operativo del piano.

Similmente a quanto fatto per il ferro da stiro, è stato applicato un sensore di temperatura in prossimità della piastra, in modo da poter segnalare, una volta terminato l'utilizzo della piastra, l'alta temperatura della stessa. Man mano che la temperatura cala e si avvicina ad una soglia di sicurezza, il segnale audio diminuisce di intensità fino a scomparire.

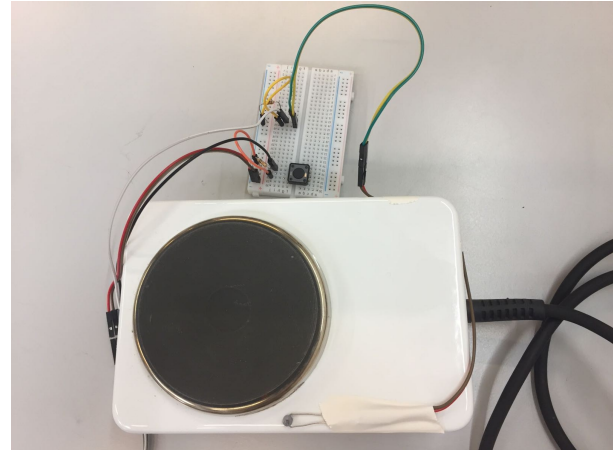


Figura 2. Piastra (elettrica) utilizzata in fase di testing.

## 4. I SUONI

La scelta dei suoni è stata effettuata cercando essenzialmente di rispettare due principi: la familiarità e la semplicità. Per familiarità intendiamo la ricerca di un suono che ricordi o sia immediatamente associabile alla situazione operativa dello strumento, sfruttando l'intuizione e l'esperienza pregressa dell'utente.

Per semplicità si intende sia la "semplicità" timbrica (infatti sono stati utilizzati solo oscillatori che producono sinusoidi e rumore bianco, e semplici effetti come filtri, phaser, riverbero, distorsione), in modo da rendere minima l'elettronica necessaria alla produzione e diffusione del suono, sia quello che riguarda la scelta delle cellule musicali utilizzate come segnale di monitoraggio, affinché l'informazione sia il più possibile intuibile ed efficace.

Al riempimento del serbatoio è stato associato un suono percussivo intonato ottenuto da un oscillatore che produce una sinusoide con intonazione di partenza intorno a 660 Hz a cui è stato applicato un filtro passa basso (con una frequenza di taglio iniziale a 800 Hz) per rendere il suono iniziale più scuro.

L'intonazione dell'oscillatore e la frequenza di taglio del filtro salgono in corrispondenza dell'aumento del livello d'acqua presente nel serbatoio, simulando quel tipico glissando naturale prodotto da un liquido che riempie un recipiente. Una volta raggiunto il livello ottimale di funzionamento, il suono di monitoraggio cessa di essere prodotto.

Il raggiungimento della temperatura ottimale di utilizzo viene segnalata da una stretta banda di rumore bianco ottenuta da un taglio passa banda fra 4.1 kHz e

6.6 kHz, a cui è stata aggiunta una distorsione che aggiunge artefatti rumorosi (tipo aliasing/downsampling) per enfatizzare maggiormente l'effetto "crepitio" tipico di un oggetto incandescente. Infine l'applicazione di un effetto phaser rende il suono più definito e con una modulazione più evidente intorno a 4.3 kHz.

Modificando i valori di questo phaser per ottenere una modulazione intorno a 6.5 kHz, abbiamo creato un suono più aspro che è stato utilizzato per segnalare l'alta temperatura del ferro.

Per l'accensione del piano di cottura è stato utilizzato un suono sinuisodale percussivo intorno a 12.8 kHz velocemente ribattuto fino all'effettiva accensione del piano. Una volta acceso viene riprodotto un suono sinusoidale a 1.3 kHz a cui è stata applicata una catena di effetti che comprende un riverbero di 1.2 secondi, un vocoder con rumore bianco come segnale portante ed un phaser che rende il suono più brillante e definito intorno a 1.1 kHz. Il suono risultante ricorda il suono del gas in uscita da un fornello a gas, ma risulta chiaro che sia di matrice elettronica.

L'alta temperatura del piano di cottura viene segnalata dal suono descritto in precedenza per la temperatura dell'acqua e del ferro da stiro, con valori leggermente diversi per quanto riguarda phaser e distorsione.

## 5. CONCLUSIONI

Il lavoro svolto dagli autori, seppur a livello di studio iniziale, ha prodotto degli elettrodomestici il cui utilizzo viene accompagnato, sostenuto e facilitato da un semplice monitoraggio audio svolto da suoni di sintesi che, in parte emulano la fisicità dell'evento ed in parte aggiungono suoni di carattere più astratto, vale a dire di cellule musicali (earcons)[4].

Nei prossimi sviluppi ci si occuperà di incorporare in modo sempre più efficace ed ergonomico la sensoristica e di rendere il sistema di sonificazione completamente integrato nei prodotti considerati.

## 6. RIFERIMENTI

- [1] Gaver, W. W. : "What in the world do we hear?" *An ecological approach to auditory source perception*. Ecological Psychology, Vol. 5, No.1, pp. 1-29, 1993.
- [2] B. Verplank: "Interaction Design Sketchbook" , 2001.
- [3] T. Hermann, A. Hunt, J. G. Neuhoff : *The Sonification Handbook*", Logos Publishing House, Berlin, 2011
- [4] Blattner, M. M., Sumikawa, D. A., & Greenberg, R. M. (1989). Earcons and icons: Their structure and common design principles. *Human-Computer Interaction*, 4(1), 11-44.

Link ad un video dimostrativo realizzato nelle fasi di test:  
<https://www.youtube.com/watch?v=23Pc6HJMBf4>

# LIS-10 E HEED: DUE APPLICAZIONI DI SHAPE MEMORY ALLOYS NELL'AMBITO DELLA SOUND ART

**Lorena Cantarut**

Conservatorio G. Tartini - Trieste  
lorena.cantarut@gmx.com

## SOMMARIO

In questo articolo viene presentata l'applicazione di una lega a memoria di forma, Shape Memory Alloy (SMA), nell'ambito della Sound Art. La principale proprietà di un SMA consiste nella capacità di richiamare una specifica forma precedentemente impressa. Nel presente lavoro, è stato utilizzato un SMA costituito da nickel e titanio, denominato Nitinol. La struttura cristallina del Nitinol può assumere due diverse fasi che dipendono dal riscaldamento della lega: a temperatura ambiente esso risulta malleabile, e quindi deformabile, mentre, se riscaldato oltre una certa temperatura, è in grado di riorganizzare la propria struttura cristallina in modo da assumere una forma prestabilita. Tale passaggio di fase è ciò che rende gli SMA ideali nell'ambito dell'interazione uomo-macchina. In questo caso la temperatura è controllata attraverso un circuito e algoritmi digitali. In particolare ho scelto di utilizzare l'analisi del suono come algoritmo. In altre parole viene fatta cambiare forma alla scultura sonora per mezzo di descrittori del suono, estratti dall'input acustico prodotto dagli utenti fruitori dell'installazione.

## 1. ROBOTIC ART E SOUND ART

Negli ultimi anni il confine tra Uomo e Macchina è andato via via assottigliandosi. Innumerevoli sono le ricerche nell'ambito dell'Intelligenza Artificiale e Machine Learning, utilizziamo elettrodomestici ed automobili sempre più intelligenti e sistemi di auto-monitoraggio delle nostre funzioni vitali tramite apparecchi indossabili. I robot sono di fatto entrati nelle nostre case. Inevitabilmente, considerate le innumerevoli possibilità apertesi, arte e tecnologia sono andate via via convergendo, dando vita a nuove discipline e ambiti di ricerca i cui confini spesso sono non ben definiti. È in questo contesto che troviamo la Robotic Art. I robot (dalla lingua Ceca "robot", lavorare) sono nati con l'intento industriale di impiegare un macchinario a sostegno dell'attività lavorativa dell'uomo. L'idea romantica di forgiare un "essere a propria immagine e somiglianza", sostenuta da una capacità di calcolo degli attuali processori sempre maggiore, ha

portato ad un tentativo di mimesi tra uomo e robot sempre più verosimile. La mimesi è effettivamente la parola chiave di questo fenomeno poiché il punto di partenza per un prototipo ha quasi sempre come riferimento il comportamento di un essere vivente. Il presente lavoro non è immune da questa tendenza. Ciò che mi ha spinto ad investigare il mondo della robotica e cibernetica, nonché l'ambito degli smart materials, è stato proprio il desiderio di sfumare ulteriormente il confine tra arte, tecnologia e natura, creando un'installazione che potesse ricordare in termini cinetici e comportamentali un essere vivente. Dagli stessi presupposti nasce anche il lavoro di Akira Nakayasu [1] il quale ha adottato la stessa tecnologia da me impiegata (Shape Memory Alloy) per creare una serie di installazioni volte ad imitare i tentacoli di un'anemone di mare ("Tentacle" 2010, "Waving Tentacles" 2014, "Luminescent Tentacles" 2016, "Tentacle Flora" 2018). I tentacoli realizzati da Nakayasu interagiscono con il movimento della mano dell'utente, muovendosi a loro volta e nei prototipi più recenti, illuminandosi. Le modalità sulle quali egli ha lavorato sono quindi la vista, sia come input che come output, e la propriocezione. Il mio lavoro, rispetto a quello di Nakayasu, si differenzia prevalentemente su questo punto: considerato il mio background in ambito acustico, mi sono soffermata su una modalità in particolare, l'ascolto, sviluppando due installazioni (LIS-10 e HEED), di cui parlerò nei prossimi paragrafi, i cui nomi rimandano chiaramente alla capacità del sistema di recepire, immagazzinare e rispondere ad input sonori. Proprio per questo motivo il mio lavoro risiede in un ambito che interseca sia la Robotic Art che la Sound Art. La Sound Art è un'espressione artistica relativamente recente, e in quanto tale non ben definita o codificata. Trova le sue origini a partire dallo sviluppo della registrazione sonora, che ha di conseguenza indotto un'attenzione nuova sulla natura del suono. Lo sviluppo delle tecnologie audio ha reso, quindi, il suono svincolato dalla sua fonte, nonché manipolabile e riproducibile come un oggetto indipendente. A partire da questo presupposto, possono essere incluse nella vasta categoria della Sound Art tutte le produzioni nell'arte contemporanea che introducono il suono come sua componente principale ed essenziale: registrazioni ambientali (field recording), sculture sonore, installazioni e performance. Un esempio di

*Copyright: © 2018 Lorena Cantarut. This is an open-access article distributed under the terms of the [Creative Commons Attribution License 3.0 Unported](https://creativecommons.org/licenses/by/3.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.*

Sound Art è riscontrabile nei lavori dell'artista Zimoun<sup>1</sup>, il quale realizza installazioni sonore a partire da "componenti semplici e funzionali", come dice lui stesso, creando pattern ritmici che sommandosi e sovrapponendosi danno vita ad un paesaggio sonoro inaspettato. Anche la mia ricerca rientra all'interno di questo settore, poiché vede il suono come protagonista, nonché rappresenta per le installazioni da me realizzate l'elemento di input, output e talvolta di entrambi.

## 2. IL NITINOL COME ATTUATORE

### 2.1 Cos'è il Nitinol?

Il Nitinol è una lega di nichel e titanio (Ni-Ti) appartenente alla famiglia dei cosiddetti Smart Materials (SM) e in particolare alla categoria dei Shape Memory Alloys (SMA). Per rientrare in questa categoria una lega deve presentare due essenziali proprietà: 1) effetto memoria di forma: ovvero la capacità di "ricordare" una particolare geometria imposta tramite alte temperature; 2) superelasticità: ovvero la capacità di subire grandi deformazioni potendo poi riacquistare la forma originale. Queste proprietà dipendono dalla capacità del SMA di assumere due diverse strutture cristalline, chiamate anche fasi: quella austenitica, stabile ad alte temperature, e quella martensitica, stabile a basse temperature. Quando la struttura è in fase martensitica i cristalli hanno la possibilità di muoversi, seppur in un solo asse. Questo permette di deformare a piacere la lega quando si trova in fase martensitica. Tuttavia, se riscaldata oltre una certa temperatura, un SMA passa alla fase austenitica, recuperando una forma precedentemente impressa.

### 2.2 SMA e Human Computer Interaction (HCI)

Grazie alle loro proprietà, all'ottima resistenza a corrosione e alla buona duttilità, gli Shape Memory Alloys (SMA) sono particolarmente adatti anche come interfacce, nelle quali possono agire da attuatori recuperando la loro forma originale, oppure esercitare forze, anche grandi, sulle strutture a cui sono vincolati. Ciò li rende notevolmente invitanti se usati come attuatori per robot [2, 3] o come display [4, 5]. Ad esempio, in [4], Velasquez et al. hanno realizzato un display dinamico composto da una matrice di pin, azionati da SMA, per caratteri Braille. Similmente gli SMA ed in generale gli SM rappresentano delle interessanti alternative come feedback aptici [6]. L'utilizzo degli SMA tuttavia non è ristretto all'ambito informatico e cibernetico, ma anzi il loro impiego è ormai una prassi consolidata, soprattutto in ambito medico ed odontoiatrico. Di leghe a memoria di forma sono ad esempio costituiti gli stent o altri divaricatori, chiodi odontoiatrici ma anche sonde guida, sistemi di autofocus, sistemi di zoom, otturatori fotografici, relè termici, valvole per rubinetteria, attuatori per climatizzatori e perfino serrature.

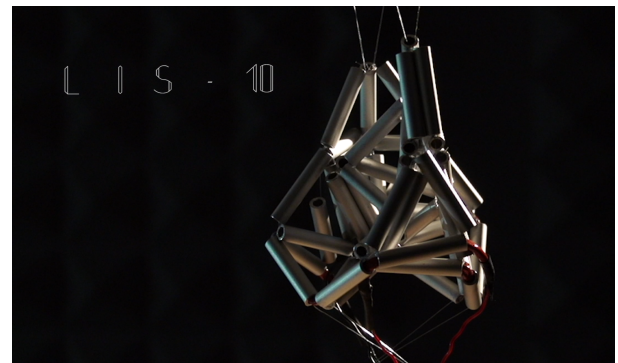
<sup>1</sup> <https://zimoun.net>

## 3. LIS-10 e HEED - SCULTURE SONORE INTERATTIVE

La mia ricerca nasce all'interno del corso di Ambienti multimodali e interattivi presso il Conservatorio G. Tartini di Trieste, con l'intento di realizzare un'installazione interattiva la quale fosse in grado di rispondere ad input esterni con un comportamento il più "organico" possibile. L'obiettivo era riuscire a modellare un materiale inanimato, rendendolo programmabile e capace di interagire con l'ambiente circostante. In altre parole mi interessava creare e sviluppare un rapporto tra artista e materia artificiale il più intimo possibile, donando a quest'ultima caratteristiche riscontrabili in un essere vivente, quali ad esempio la capacità di ascoltare e reagire di conseguenza. Al di là del primo intento, si sono subito aperte altre possibili strade per LIS-10 e HEED. Una di queste è di certo la capacità delle sculture di agire da performer, con o senza la presenza di un essere umano. Un riferimento è riscontrabile nel lavoro di Bill Vorn e della sua Robotic Art. In particolare parlo di "Grace State Machine"<sup>2</sup>, una performance che vede come protagonisti un essere umano e due macchine, investigando il rapporto tra due diversi tipi di corpi, organico e sintetico, e le possibilità espressive di essi sia separatamente che come unico grande organismo.

### 3.1 LIS-10, Primo Prototipo

Il primo prototipo di scultura è stato "LIS-10, a Living Interactive Sculpture" (Figura 1). Il nome è anche un gioco di parole che rimanda al verbo "Listen", suggerendo la natura sonora dell'installazione.

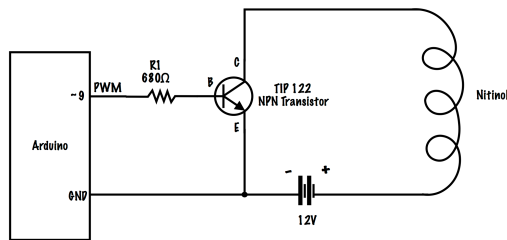


**Figura 1.** LIS-10, primo prototipo di scultura sonora utilizzando il Nitinol come attuttore.

LIS-10 è composto da una struttura modulare, che se chiusa su sé stessa raggiunge le dimensioni di circa 13cm x 9cm x 9cm. Essa è formata da tubicini di alluminio all'interno dei quali passano due fili di Nitinol, ricoperti da una guaina isolante e termo-resistente. In fase martensitica i fili assumono la forma della geometria imposta dalla struttura modulare, mentre in fase austenitica essi

<sup>2</sup> <http://www.billvorn.concordia.ca/menuall.html>

tendono ad assumere una forma precedentemente impressa (fili distesi) che va quindi a deformare la struttura modulare, espandendola orizzontalmente. La fase austenitica è ottenuta riscaldando i fili di Nitinol attraverso un circuito, nel quale il voltaggio è controllato da pulse-width modulation (Figura 2). Eliminando il voltaggio passante attraverso il circuito, i fili, raffreddandosi, ritornano in fase martensitica, diventando nuovamente soggetti a deformazione, per azione della forza di gravità amplificata dalla struttura in alluminio: questo causa la contrazione della scultura.



**Figura 2.** Circuito utilizzato per attivare il passaggio di fase di un filo di Nitinol.

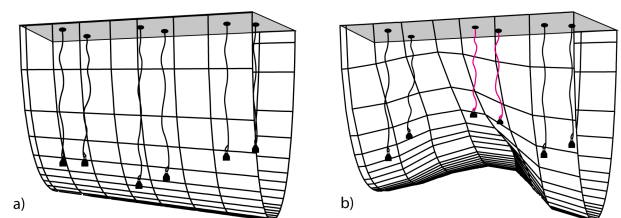
Il suono ha una duplice funzione: input e output. Infatti, il segnale di controllo per il circuito è l'ampiezza del suono prodotto dall'utente (input), catturato da un microfono dinamico posizionato a fianco alla scultura. Quando l'ampiezza supera la soglia di attivazione un algoritmo, sviluppato all'interno dell'ambiente Max/MSP, riscalda l'ampiezza in un range da 0 a 255, il quale rappresenta il segnale di controllo della tensione inviata ad Arduino tramite trasmissione seriale. Per evitare surriscaldamenti ho inserito un blocco automatico, il quale interrompe il flusso di dati inviati ad Arduino nel caso in cui la durata della trasmissione sia maggiore di 4 secondi. Al fine di migliorare il raffreddamento dei fili di Nitinol, e aumentarne quindi il potenziale cinetico, ho implementato un meccanismo di rotazione nell'invio di tensione: ogni volta che l'ampiezza dell'input supera la soglia di attivazione, viene riscaldato un filo diverso. Espansione e contrazione della scultura sono sonificati (output) anch'essi all'interno di Max/MSP, in modo da tradurre il movimento in un dominio acustico, rendendo udibile il passaggio dalla fase martensitica a quella austenitica e viceversa. Il segnale di controllo della tensione è quindi anche segnale di controllo per l'ampiezza dell'output acustico. La sonificazione è utile anche in un'ottica di interazione con la scultura perché, restituendo un feedback sulla fase in cui essa si trova, permette all'utente di capire quando è nuovamente possibile fornire un nuovo input. A tal fine ho utilizzato dei campioni, uno per ogni filo, che fossero in grado di mimare verosimilmente il movimento compiuto nel momento dell'attivazione, aggiungendo in testa ed in coda un attacco e decadimento piuttosto lenti in modo da rispecchiare anche la rallentata capacità del materiale di scaldarsi e raffreddarsi. Durante l'interazione con il pubblico, di cui parlerò in maniera più

estesa nel prossimo capitolo, tale accorgimento è stato molto importante poiché, non essendo la reazione cinetica immediata rispetto all'input prodotto dagli utenti, una risposta acustica istantanea assicurava all'utente che il suono emesso era stato riconosciuto dalla scultura. Il principale problema dei SMA è dovuto all'isteresi che presentano questo tipo di materiali, e quindi all'asimmetria tra la velocità di riscaldamento e raffreddamento. Questo rappresenta un grosso limite per la qualità della risposta e dell'interazione poiché è necessario un determinato lasso temporale tra un input e quello successivo. Un video dimostrativo della prima versione di LIS-10 è disponibile al seguente indirizzo:

<https://vimeo.com/273961256>

### 3.2 HEED, Secondo Prototipo

HEED, dall'inglese "prestare attenzione", nasce allo scopo di migliorare l'interazione utente-scultura, mediante una più raffinata analisi dell'input acustico fornito dall'utente ed una più articolata mobilità della scultura. HEED è composto da sei fili di Nitinol di circa 30cm, appesi verticalmente ad un supporto rettangolare delle dimensioni di 34cm x 67cm. L'estremità inferiore di ogni filo è fissata ad un tessuto semi-trasparente a cui sono attaccati dei fogli di alluminio. In fase martensitica i fili si presentano distesi, mentre in fase austenitica assumono una forma elicoidale contratta, poiché questa è la forma che ho impresso al Nitinol (Figura 3). In questo modo il passaggio di fase causa una deformazione localizzata del tessuto, nel punto in cui è fissato ciascun filo di Nitinol. In HEED, non essendo i fili inseriti all'interno di una struttura rigida e potendosi quindi muovere liberamente, l'ampiezza del movimento di ogni filo è maggiore rispetto a LIS-10. Questo si traduce in una migliorata mobilità della scultura. Inoltre, a differenza del primo prototipo nel quale erano impiegati due fili, in HEED l'utilizzo di sei fili ha reso possibile una migliore distribuzione del movimento, nonché una migliore gestione del raffreddamento di ogni filo. Il ritorno alla forma distesa, in fase martensitica, è accelerato dall'aggiunta di un peso (30grammi circa) all'estremità inferiore del filo.



**Figura 3.** Illustrazione di HEED: a) modalità a riposo, b) con i due fili centrali (indicati in fucsia) in azione.

Il secondo intervento si è focalizzato su una miglior analisi dell'input acustico, tramite l'estrazione di descrittori del suono, quali *brightness* e *noisiness*. La velocità del movimento è variata in base alla *brightness*, mentre alla *noisiness* corrisponde un pattern di movimento più o meno aleatorio. Per l'estrazione di tali informazioni si è scelto di utilizzare l'oggetto *analyzer*<sup>3</sup>, all'interno di Max/MSP. Anche in questo caso il suono ha una duplice funzione di input e output. Tuttavia, alla sonificazione, ho preferito l'utilizzo del suono prodotto dalla scultura stessa. L'impiego di fogli di alluminio all'interno della scultura ha, appunto, lo scopo di restituire un feedback acustico della deformazione della scultura. Ho scelto, infine, di impiegare un Electroluminescent wire, controllato anch'esso tramite pulse-width modulation, per notificare all'utente lo stato di attivazione o disattivazione della scultura: in questo caso il parametro di controllo è la *loudness* del suono, estratta anch'essa attraverso l'oggetto *analyzer*.

#### 4. FEEDBACK DAL PUBBLICO

Durante il percorso di ricerca ho avuto modo di ricevere preziosi feedback dal pubblico grazie a diverse esposizioni. Esse riguardano principalmente LIS-10. LIS-10 è stata presentata al pubblico per la prima volta come installazione durante "La Settimana della Composizione" presso il Conservatorio "G. Tartini" di Trieste e, a distanza di una settimana, durante l'Open Day sempre presso il Conservatorio. Il pubblico che ha potuto usufruire dell'installazione era molto diverso tra un evento e l'altro e ciò mi ha permesso di avere un campione di utenti molto vasto ed eterogeneo. Nel primo evento gli utenti avevano alle spalle un background artistico e musicale, mentre nel secondo evento gli utenti erano principalmente bambini e ragazzi accompagnati dai genitori. La reazione osservata tuttavia è stata molto simile: un'iniziale senso di disagio nei confronti della scultura sonora nel momento in cui iniziava a muoversi. Il motivo, stando al commento più frequente, è dovuto alla discrasia tra il movimento, così simile a quello di un essere vivente, e l'aspetto asettico e apparentemente inanimato dell'alluminio che compone la scultura.

#### 5. SVILUPPI FUTURI

La principale caratteristica di LIS-10 e HEED è quella di essere sistemi capaci di ascoltare e rispondere agli input acustici prodotti dagli utenti. In altre parole sono sistemi interattivi. Tuttavia le capacità di riconoscimento e di risposta, nonostante i risultati ottenuti con il secondo prototipo, HEED, hanno ancora ampio margine di miglioramento. Per quanto riguarda la parte di riconoscimento è necessario uno sviluppo sul lato della programmazione, mentre una più accurata risposta del movimento è da ricercare tramite un ulteriore studio e ricerca nell'ambito

dei materiali da utilizzare in sinergia con il Nitinol e nella complessiva gestione cinetica della scultura. Tali miglioramenti si tradurrebbero quindi in una maggiore potenzialità espressiva, che renderebbe le sculture molto più efficaci sia come agenti interattivi che come possibili performer.

#### 6. BIBLIOGRAFIA

- [1] A. Nakayasu: "Waving Tentacles: A System and Method for Controlling a SMA Actuator," *SIGGRAPH 2014 Posters*, ACM, 2014.
- [2] M. Sreekumar, T. Nagarajan, M. Singaperumal, M. Zoppi, R. Molfino, "Critical review of current trends in shape memory alloy actuators for intelligent robots", *Industrial Robot: An International Journal*, Vol. 34, pp.285-294, 2007.
- [3] J.K. Paik and R.J. Wood: "A bidirectional shape memory alloy folding actuator," *Smart Materials and Structures*, Vol. 21, No. 6, 2012.
- [4] R. Velazquez, E.E. Pissaloux, and M. Wiertelwski: "A Compact Tactile Display for the Blind with Shape Memory Alloys," *Proceedings of the 2006 IEEE International Conference on Robotics and Automation*, pp. 3905-3910, 2006.
- [5] R. Vitushinsky, S. Schmitz, and A. Ludwig: "Bistable Thin-Film Shape Memory Actuators for Applications in Tactile Displays," *Journal of microelectromechanical systems*, Vol. 18, No. 1, pp. 186-194, 2009.
- [6] D. Hwang, J. Lee, and K. Kim: "On the design of a miniature haptic ring for cutaneous force feedback using shape memory alloy actuators," *Smart Materials and Structures*, Vol. 26, No.10, 2017.

<sup>3</sup> <http://cnmat.berkeley.edu/downloads/>

# MYTRENDS: PLAYING WORDS

Michela Coppola

Conservatorio di Musica “Giuseppe Verdi  
di Milano”

micopp77@yahoo.it

## ABSTRACT

MyTrends è una ‘macchina musicale’ interattiva, capace di generare una composizione. L’obiettivo è quello di rendere possibile, anche per coloro che non hanno specifiche competenze, la creazione di un brano musicale, mediante un sistema semplice e intuitivo. Nella lingua inglese il verbo play, come il francese jouer, indica le azioni del ‘suonare’ e ‘giocare’: il progetto si ispira alle pratiche musicali e agli antichi passatempi diffusi nell’Europa del XVIII e XIX secolo, come i giochi musicali dei dadi. Per favorire un approccio immediato e intuitivo dell’utente, si è deciso di adoperare uno strumento diffuso nella società contemporanea come Internet. MyTrends chiede a chi lo utilizza di compiere alcune operazioni quotidiane: sedersi davanti ad una tastiera e ad un monitor su cui comparirà la familiare grafica di Google ed effettuare la ricerca di una parola. MyTrends ‘interroga’ il sistema di report Google Trends e interpreta i dati rilevati, elaborando l’algoritmo di sonificazione della parola digitata. Il sistema consente all’utente di cimentarsi anche nelle fasi di interpretazione e variazione del materiale sonoro generato dalla macchina, mediante interazione via mouse.

## 1. INTRODUZIONE

« *Management Style is a replication of patterning ... An individual’s style of speaking or writing, for instance, results in large part from lexical, grammatical and syntactic choices made within the constraints of the language and dialect he has learned to use but does not himself create. And so it is in music, paintings and the other arts. [1]* »

Il progetto ha l’obiettivo di creare una macchina interattiva capace di elaborare le parole digitate sul web al fine di consentire, anche a coloro i quali non hanno competenze di teoria musicale, di realizzare la sonificazione dei dati rilevati. Questa tematica ha un suo antecedente storico rappresentato dal gioco musicale dei dadi, passatempo popolare nell’Europa del XVIII e XIX secolo. Pare che anche Wolfgang Amadeus Mozart si sia cimentato nella creazione di giochi musicali che consentivano di comporre minuetti «senza la minima conoscenza della musica».

Copyright: © 2018 Michela Coppola. This is an open-access article distributed under the terms of the [Creative Commons Attribution License 3.0 Unported](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

In accordo con la definizione di Gerhard Haupenthal [2], possiamo definire il gioco musicale dei dadi un’architettura creata attraverso la combinazione di una serie di strutture musicali formulate da un autore, la cui successione viene organizzata secondo criteri casuali in modo da creare pezzi musicali, la cui forma è definita indipendentemente dalla selezione dei singoli elementi. Ne risulta un processo compositivo meccanizzato, capace di organizzare i suoni in base ai dati ricevuti, che ricorda quasi una macchina pensante, in analogia con i procedimenti della composizione algoritmica.

Molte sono state le rivisitazioni dell’esperimento mozartiano mediante l’applicazione delle tecniche della composizione algoritmica. Alcuni esempi più recenti:

- *Mozart*, progetto realizzato da Marian Aldenhövel basata su browser con partitura e generatore MIDI, in grado di generare una partitura tramite un testo o nome arbitrario, o tirando i dadi, nelle due modalità download e playback della partitura<sup>1</sup>.

- *Mozart Dice Game* di Konstantin Weixelbaum, una App disponibile su App Store per dispositivi iOS<sup>2</sup>.

- *Musikalische Würfelspiele*, di Peter Baumann, nella nuova versione 2.0 del 2007<sup>3</sup>, è un programma per creare brani musicali, basati sul principio dei giochi di dadi musicali, suonandoli con uno dei 128 strumenti MIDI disponibili

- *Musikalische Würfelspiele* programmato in Python da Moisés Cachay<sup>4</sup>. Lo script, originariamente concepito per generare i brani in un account Twitter (@wurfelspielbot), può essere utilizzato in generale per la creazione di spartiti e file audio.

MyTrends mostra i risultati di una ricerca per l’utilizzo dell’interfaccia gestuale che consente la riproduzione e il controllo di eventi sonori in tempo reale, con l’intento di valorizzare il contributo del gesto quale strumento di pratica e organizzazione musicale, attra-

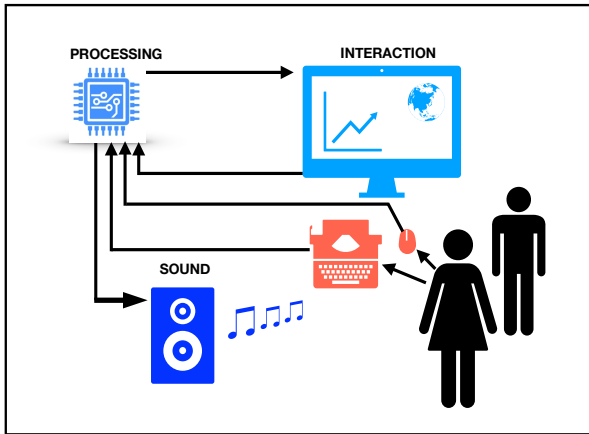
<sup>1</sup> <https://mozart.qvwx.de/index.en.html>

<sup>2</sup> <https://itunes.apple.com/us/app/mozart-dice-game/id946580946?l=de&ls=1&mt=8>

<sup>3</sup> Per i dettagli su Google Trends si rimanda al paragrafo 1.1 <https://www.combib.de/programme/musikalischewuerfelspiele.html#Programm>

<sup>4</sup> <https://github.com/Xpktro/wurfelspiel>

verso l'impiego della tecnologia informatica. L'obiettivo di questo sistema è quello di fornire sia all'utente sia agli ascoltatori (potenziali attori di nuove interpretazioni) un approccio di tipo 'strumentale' espressivo al suono, mediante l'impiego di relazioni semplici e intuitive, utili anche alla memorizzazione delle diverse tipologie di gesture e ai risultati sonori ad esse corrispondenti. Nella figura 1 sono mostrate le fasi dell'installazione. L'utente innesca il processo di sonificazione digitando la sua ricerca su Google Trends, un servizio di analisi che consente di conoscere, in risposta alle richieste dell'utente, la frequenza di ricerca di una specifica parola o frase sul web.



**Figura 1.** Elementi costitutivi dell'installazione MyTrends

Il risultato della ricerca effettuata è proiettato in tempo reale attraverso la schermata del grafico Google Trends, che sintetizza il processo di sonificazione dei dati. Il mapping del grafico estrae i dati necessari per la sonificazione e il risultato sonoro generato dalla macchina è suscettibile di modifiche da parte dell'utente. Il grafico di Google Trends assume, dunque, il ruolo di una 'partitura', che viene interpretata tramite i movimenti del mouse dagli utenti.

### 1.1 Big Data e Google Trends

Al giorno d'oggi, gli algoritmi sono in grado di trattare in breve tempo molte variabili gestendo una grande quantità di informazioni. Questa grande rivoluzione è stata inglobata nella logica del marketing, dove l'uso dei Big Data tende a costruire così detti 'metodi di raccomandazione', utilizzati per elaborare proposte pubblicitarie formulate sulla base degli interessi specifici del cliente. Gli algoritmi per il tracciamento delle ricerche sul web riescono a predire situazioni relative alla vita degli utenti per utilizzarle in funzione degli scopi più disparati. Il cuore del progetto MyTrends è quello di utilizzare i Big Data per creare musica.

Nello specifico, Google Trends è un servizio di analisi delle tendenze di ricerca sul web utilizzabile per effettuare analisi di mercato; le tendenze correnti (trend) vengono classificate in ordine di frequenza di ricerca. I risultati sono mostrati attraverso un grafico che sintetizza l'andamento della popolarità della ricerca o vi-

sualizzazione di una data parola nel tempo. L'intervallo di tempo è personalizzabile ed è estendibile dalla ricerca di dati in un intervallo che va dall'ultima ora fino all'anno 2004. La ricerca e la visualizzazione sono impostabili per nazione e per lingua. Il servizio è raggiungibile da tutti all'indirizzo: <https://trends.google.com/trends/>.

### 1.2 L'estetica dell'installazione

La ricerca estetica in MyTrends parte dal presupposto che la scoperta delle sonorità della voce possa essere considerata l'espressione più antica del rapporto uomo-musica. La parola è da sempre in relazione con il suono, questo si evince dalle sue diverse destinazioni che spaziano dalla musicalità della poesia alle teorie psicoacustiche di Diana Deutsch [5]; si pensi anche ad alcune composizioni elettroacustiche, come *Different Trains* di Steve Reich o alle composizioni per voce di Berio, solo per citare alcuni esempi. MyTrends non utilizza le parole secondo caratteristiche strumentali legate alle funzioni comunicative del linguaggio, ma scompone la parola, per lavorare sul suono puro, sul ritmo, sul timbro, sulle sue libere risonanze. Ogni lettera rappresenta quindi un oggetto sonoro, inteso in senso schaefferiano, ovvero nella sua natura morfologica [3,4]. L'estetica di MyTrends nasce da una concezione olistica del progetto installativo, ispirata al modello descritto da Hunt e Wanderley [6]. L'obiettivo è quello di 'restituire' all'utente l'unicità della parola digitata, superando i sistemi di calcolo che regolano i Big Data. L'interazione gioca un ruolo fondamentale nel processo poetico, come metafora della possibilità di interpretare e deviare il flusso dei dati. L'utente è in grado, anche senza aver generato una propria partitura, di intervenire sulla sonificazione. Questa modalità è già stata realizzata in passato e ne è un esempio il progetto Ariadne [7], in cui l'interazione utente-macchina valorizza l'unicità del momento performativo, centrale nell'esecuzione di un qualsiasi brano musicale. In MyTrends attraverso l'esperienza personale, l'utente può di volta in volta interpretare la 'partitura' (il grafico).

## 2. SONIFICAZIONE

### 2.1 Processi

Il lavoro di scomposizione e modifica effettuato sulle parole digitate rappresenta il cuore di MyTrends. In figura 2 viene mostrato il diagramma rappresentativo dei processi, sintetizzabili in tre fasi:

- Trasformazione del file di testo in file audio. Il sistema utilizza gli strumenti Text to Speech interni a Mac OSX: tramite l'oggetto Shell in Max/MSP, che attua il collegamento con il terminale di sistema, è possibile convertire qualsiasi file .RTF o .TXT in un file audio, salvarlo in un buffer di memoria per il processing.

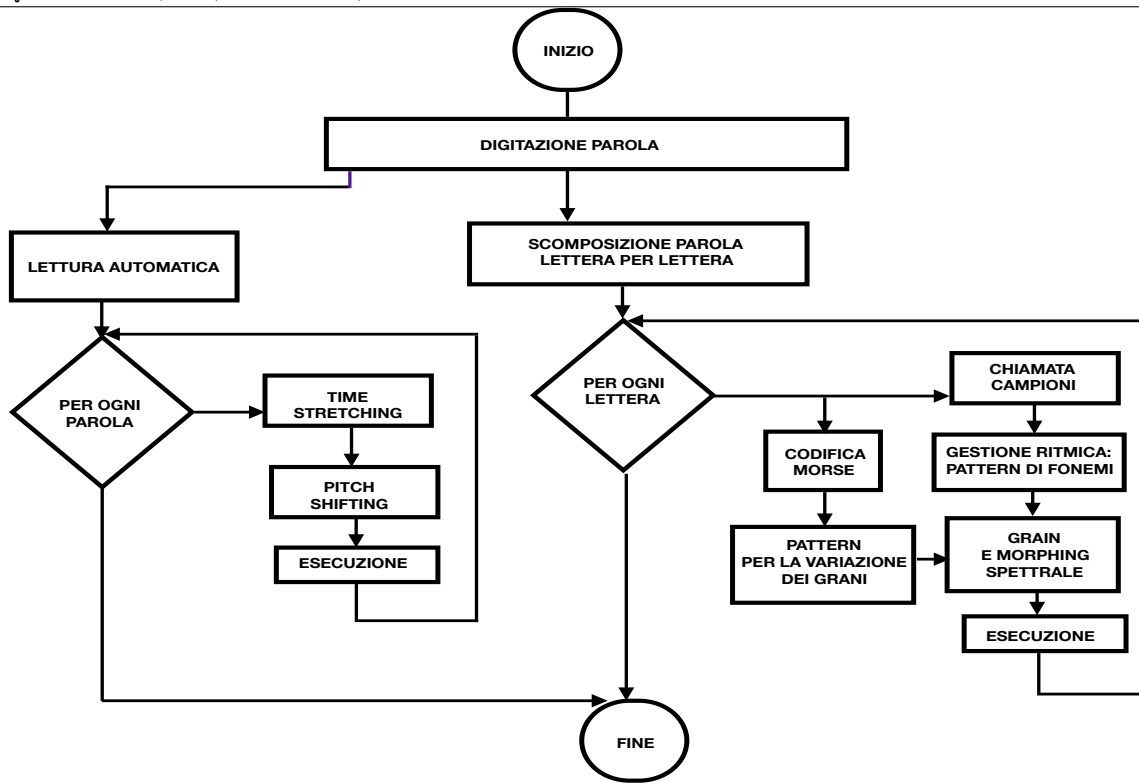


Figura 2. Diagramma rappresentativo dei processi

- Isolamento, via Max/MSP, delle singole lettere che compongono la parola.
- Codifica della query in Codice Morse per l’attivazione dei processi di granularizzazione sui campioni.

Il risultato della digitazione è innanzitutto eseguito come file audio, allo scopo di sottolineare la relazione tra la parola digitata e il risultato della sonificazione. Questo risultato è successivamente esposto al processing audio. La parola viene inoltre scomposta in lettere singole, che attivano i campioni audio memorizzati in un database integrato in Max/MSP. Il sistema codifica le lettere secondo il Codice Morse generando una linea ritmica tipica del sistema morse, utilizzata per l’attivazione di processi di granularizzazione effettuati sui campioni ricavati dagli strumenti Text to Speech interni a Mac OSX.

## 2.2 Materiali utilizzati

I materiali utilizzati sono:

- suoni sintetici generati dagli strumenti Text to Speech interni a Mac OSX.
- suoni naturali ricavati mediante sessione di registrazione di cantanti lirici, cui è stato chiesto di effettuare improvvisazioni vocali sui fonemi della lingua italiana.

La sessione di registrazione si è svolta presso il Conservatorio “Giuseppe Verdi” di Milano ed ha coinvolto voci femminili e maschili (estensione: soprano e tenore), per assicurare la varietà timbrica dei campioni. Il materiale ricavato dalle improvvisazioni vocali è utilizzato per la sonorizzazione delle singole lettere in cui è scomposta la parola. Per ogni lettera sono stati regi-

strati da un minimo di cinque a un massimo di diciotto campioni, in modo da avere materiali sonori diversificati per ogni digitazione. I suoni naturali sono selezionati di volta in volta dalla macchina secondo un sistema di chiamata random applicato all’oggetto coll di Max/MSP. Tutti i suoni (sintetici e naturali) sono processati in tempo reale così da generare fasce e suoni puntillistici.

## 2.3 Principi della sonificazione e algoritmo compositivo

Il sistema interpreta le lettere digitate in base alla codifica di caratteri nota come American Standard Code for Information Interchange (ASCII). Le digitazioni dell’utente vengono interpretate mediante decodifica dei 95 caratteri ASCII stampabili, numerati da 32 a 126. Il risultato viene mostrato nel browser all’interno di un frame Javascript in Max/MSP.

L’algoritmo compositivo di MyTrends è ispirato alle regole del gioco mozartiano, che secondo alcune ipotesi mostrano il tentativo del compositore di tradurre in musica un gioco di combinazioni sul nome della sorella Francisca<sup>5</sup>. Si elencano di seguito le cinque regole probabilmente utilizzate dall’autore:

1. Scrivi un nome aggiungendo la lettera “z” alla fine/
2. Riscrivi le lettere in ordine alfabetico/
3. Se una lettera ricorre, posizionala nella seconda apparizione alla fine del gruppo principale e separata dal segno +/-
4. Assegna il numero 1 o 2 a ciascuna lettera una dopo l’altra; dove il numero 1 è assegnato al gruppo principale, il numero 2 deve essere assegnato al gruppo dopo il segno +, e viceversa, per non duplicare numeri e

<sup>5</sup>

Si veda: <http://www.asahi-net.or.jp/~rb5h-ngc/e/k516f.htm>

quindi figure musicali/ 5. Torna all'ortografia originale.

Le regole del gioco mozartiano sono state realizzate in Max/MSP utilizzando l'oggetto zl che consente di compiere operazioni sulle liste, operazioni possibili una volta scomposta la parola nella lista delle singole lettere. L'algoritmo compositivo di MyTrends si basa sulle prime quattro regole del gioco e utilizza alcune funzioni di zl, disponibili impostando il primo argomento nell'oggetto. Le modalità di funzionamento di zl utilizzate per realizzare le regole del gioco sono: zl join - che consente di unire le liste (regola n.1), zl sort - che consente di ordinare un elenco di input in ordine crescente o decrescente (regola n.2), zl.group - che specifica il numero degli elementi della lista memorizzati più di recente (regola n.3). Per la regola n.3 è possibile dividere le liste utilizzando l'oggetto iter, che spacchetta e mostra il contenuto della lista per singoli elementi, memorizzandolo in un file di testo mediante l'oggetto coll. Questa operazione è necessaria e consente di dividere la lista in ingresso in modo da separare il testo senza duplicazioni delle lettere dalle ripetizioni dei caratteri alfabetici. È poi possibile tornare alla gestione delle singole lettere tramite l'oggetto unpack e indirizzarle allo Step Sequencer per la gestione dei campioni audio. La regola n.4 è realizzata in due sezioni diversificate: una per il testo senza duplicazioni, una per le ripetizioni delle lettere. Le due sezioni sono organizzate in modo da garantire che scelta di uno stesso campione audio non si ripeta: in caso di duplicazioni di caratteri il sistema è in grado di selezionare campioni sempre diversi.

La gestione ritmica del materiale audio viene dedotta dal calcolo della velocità delle digitazioni dell'utente. Il sistema software analizza e misura la distanza tem-

porale tra le ultime lettere digitate, individuando i battiti per minuto (BPM). Il calcolo del BPM viene riportato in millisecondi in base all'unità di movimento ed è ricavato dalla differenza temporale  $\Delta T$  tra una digitazione e l'altra:

$$\Delta T = Tap(x) - Tap(y) \tag{1}$$

Calcolando la distanza tra le digitazioni in millisecondi, avremo, dunque, i valori del BPM:

$$BPM = 60000(ms) / \Delta T \tag{2}$$

### 3. IMPLEMENTAZIONI

MyTrends si avvale di dispositivi software e hardware e consente una forte interazione umana nei processi di gestione del suono e di creazione musicale. Come mostrato in figura 3, per la ricerca delle query, la proiezione dei dati via web browser e il trattamento audio viene utilizzato il software Max/MSP, con l'oggetto jweb, in Javascript, ottimizzato per utilizzare il browser web. Per l'estrapolazione e l'elaborazione dei dati è stato creato uno script Python che impiega una chiave-sviluppatore per poter utilizzare le API di Google Trends. La comunicazione tra Python e Max/MSP è gestita tramite il protocollo di eventi discreti Open Sound Control (OSC) che consente la comunicazione tra controller, computer, sintetizzatori audio e altri dispositivi multimediali ottimizzati per la tecnologia di rete, come descritto da Wessel e Wright [8]. Le lettere digitate, una volta instradata la query per la proiezione del grafico via Max/MSP, vengono codificate in messaggi OSC e sono trasmesse al codice Python per la raccolta dei dati, necessaria alla gestione dell'algoritmo compositivo.

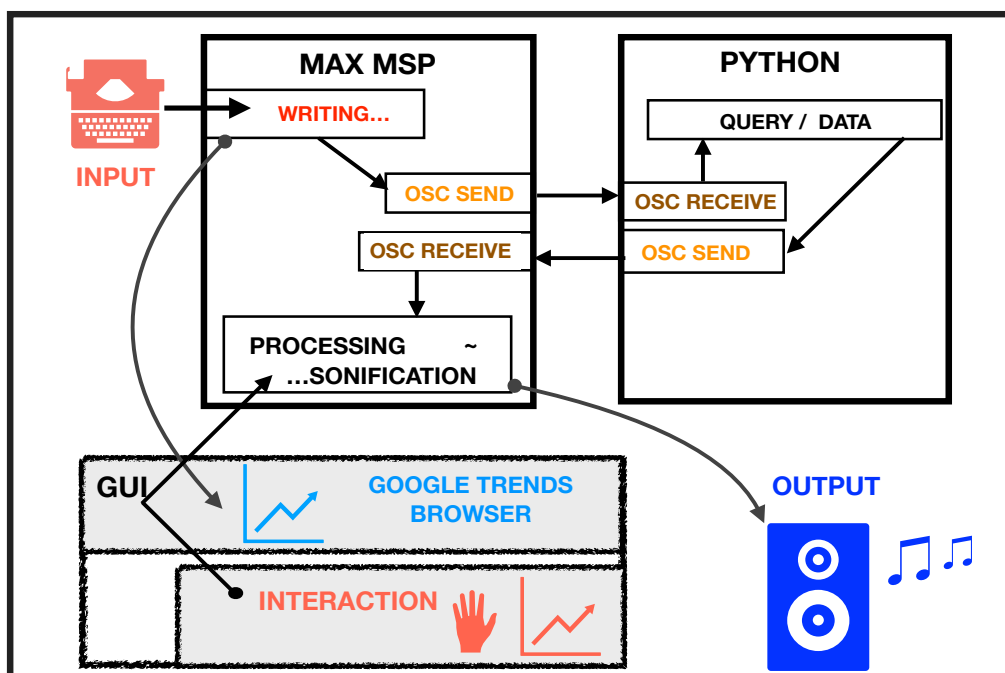


Figura 3. Schema delle comunicazioni tra Max/MSP, Python e l'interfaccia grafica

## 4. INTERAZIONE

In relazione alla fruibilità del dispositivo, ci si è posto il problema dell'accessibilità, della sperimentabilità, delle possibilità di controllo delle caratteristiche e della gestione del tempo sulla scia di quanto evidenziato da Wanderley e Orio [9]. Si parte dal presupposto che il ruolo del gesto sia fondamentale nella pratica esecutiva, così come in molte pratiche della nostra vita. Tramite le scienze psico-pedagogiche è stato dimostrato come il corpo sia uno strumento centrale della nostra conoscenza, sia intellettuale che pratica, come evidenziato da Klemmer, Hartmann e Takayama [10]. Le interfacce che implicano il coinvolgimento del corpo possono sfruttare la cognizione fisica come vantaggio per l'apprendimento, anche in termini temporali. MyTrends offre all'utente una modalità di interazione semplice e diretta, per dare l'opportunità, anche a chi non è esperto in ambito performativo, di sperimentare le caratteristiche dell'espressività musicale. Una volta attivata la sonificazione, le modifiche apportate in modo intuitivo dai gesti degli utenti sul materiale sonoro prodotto dalla macchina: timbro, dinamiche, altezze sono manipolate di volta in volta tramite mouse, con risultati diversi a seconda che l'utente decida di seguire l'andamento del grafico o di variarne le caratteristiche. Qui le tecniche di 'thinking through doing' e 'learning through doing' rappresentano il nucleo di una sperimentazione personalizzata promossa come approccio per la conoscenza e la gestione dei parametri del suono.

### 4.1 Una macchina interattiva: interpretazione della sonificazione e performance

La computer music interattiva è ormai un campo altamente specializzato di Interazione uomo-computer (Human Computer Interaction, HCI), poiché coinvolge diverse abilità cognitive e motorie complesse, come esplicitato nell'articolo di Wanderley e Orio [9]. Una caratteristica importante dei sistemi HCI risiede nel fatto che l'obiettivo dell'interazione - la performance - si basa sulla comunicazione bidirezionale tra esecutore e computer.

I gesti del performer rappresentano una scelta dell'utente, ma sono anche valori di input per il sistema; così come, allo stesso tempo, il risultato della sonificazione viene ascoltato dal performer, che può utilizzarlo per estrarre informazioni sullo stato del sistema stesso. L'interazione con MyTrends ha modalità di tipo esplorativo: l'utente scopre da solo come controllare il dispositivo, sperimentando le diverse combinazioni di controllo di input, con una risposta immediata del sistema.

### 4.2 Mapping

Il mapping è spesso usato, in relazione alla composizione algoritmica, quando un parametro che ha determinati valori viene dimensionato o trasformato in modo che possa essere usato per controllare un altro parametro o un insieme di parametri, quali ad esempio

quelli di un motore di sintesi, con riferimento alla trattazione di Hunt e Wanderley [6].

Per quanto riguarda l'interpretazione dei Big Data, i valori si presentano dimensionati nel grafico di Google Trends in un range che va da 0 a 100, sull'asse delle ordinate. I dati vengono utilizzati da MyTrends come parametri di inviluppo della composizione: le liste di valori, estratte dallo script Python ad ogni ricerca digitata, sono raccolte in Max/Msp mediante multislider che le trasmette, di volta in volta, come liste di ampiezze al convertitore digitale analogico (DAC). I dati raccolti sono utilizzati, inoltre, per controllare:

- Banco di Filtri realizzato in Max/Msp mediante Fast Fourier Transform (FFT): il segnale in ingresso viene diviso in bande mediante FFT; ciascun bin di frequenza viene moltiplicato per ognuno dei valori ricavati dallo script Python. Il codice prevede una serie di riletture di modo tale che l'indice FFT di  $fftin\sim$  sia riletto costantemente, mediante la lista generata da un multislider che raccoglie i dati estratti da Python.

- Pitch Shifter per le variazioni di frequenza 'suoni naturali' (si veda il paragrafo 2.2) ad ogni operazione di riletture del campione. In questo caso i valori estratti dallo script Python determinano variazioni di frequenza nel tempo.

MyTrends utilizza il mouse come dispositivo di input per il controllo di alcuni parametri. Il sistema acquisisce le variazioni delle posizioni x, y e button state. Pitch e panning dei campioni sono controllati, dunque, dai valori del mouse. Il controllo può essere continuo - nel caso di gesti di modulazione e modifica - o discontinuo nel caso di gesti di selezione da parte dell'utente.

Le strategie di mapping utilizzate sono di tipo:

- *uno-a-uno* (un parametro di sintesi è guidato da un parametro performativo), come avviene per l'attivazione del volume tramite l'oggetto mousestate di Max/MSP. Una volta attivato il controllo del mouse, l'audio sarà attivato quando vengono registrati i movimenti della mano, in modo tale che l'energia trasmessa dall'utente sia direttamente responsabile del suono, come raccomandato da Manzo [11]; questo consente all'utente di acquisire la consapevolezza della stretta relazione causa-effetto tra suono e gesto, come si evince dal lavoro di Hunt, Wanderley e Paradis [12];

- *uno-a-molti* (un parametro performativo influisce su diversi parametri di sintesi), come per la digitazione delle singole lettere che influenzano sia l'aspetto ritmico che l'aspetto timbrico della composizione, nonché le variazioni del processing.

MyTrends propone, dunque, una forma di interazione di tipo multiparametrico e performativo, adatta al controllo in tempo reale, mediante cui l'utente possa esplorare le caratteristiche espressive del suono, in linea con lo studio di Hunt e Kirk [13].

## 5. VALUTAZIONI E SVILUPPI

MyTrends rappresenta il primo passo di una ricerca personale nata da una riflessione sul sistema di manipolazione dei dati attraverso l'uso del web e dei social media. Questa ricerca, mossa da un interesse personale per la composizione algoritmica interattiva, ha generato l'idea di realizzare una macchina che consenta anche a un neofita della composizione di cimentarsi nella creazione musicale. Con il primo esperimento è stato registrato un bilancio positivo dal punto di vista artistico, sollecitando largo interesse su un bacino d'utenza eterogeneo dal punto di vista culturale, composto da utenti provenienti da aree geografiche diverse. Il sistema è stato testato presso l'Istituto Professionale Socio-Sanitario Oriani Mazzini di Milano, da 15 studenti - molti dei quali stranieri con scarsa conoscenza della lingua italiana - iscritti al secondo anno di corso, durante l'ora di laboratorio musicale. Gli utenti avevano una formazione musicale che corrisponde alle conoscenze musicali di base previste dai programmi ministeriali per il terzo anno della scuola secondaria di primo grado e non si erano mai cimentati con la composizione, tantomeno con la pratica strumentale. Un bilancio positivo riguarda sicuramente l'approccio interattivo. La scelta di strumenti quali mouse e tastiera ha contribuito in maniera positiva all'interazione. Ci si è resi conto del grado di personalizzazione che offre il sistema sul piano cognitivo. Sperimentando MyTrends, l'utente diviene di fatto esecutore della sonificazione e si percepisce come soggetto attivo; questo ha stimolato l'utente a condividere emozioni positive con il gruppo di uditori. La statistica sul gradimento di MyTrends ha registrato buoni risultati. L'interazione con MyTrends ha stimolato l'attitudine all'attenzione e all'aggregazione tra gli utenti. Il bilancio positivo deve tener conto, però, della fascia d'età degli utenti: pur trattandosi di neofiti della musica, tutti i soggetti mostravano una buona propensione all'uso della tecnologia senza impedimenti di tipo operativo-pratico, ma sarebbe auspicabile poter ripetere l'esperimento con un bacino d'utenza di età più matura e meno avvezzata all'utilizzo della tecnologia. L'esperimento era volto a verificare la funzionalità dell'interazione uomo-macchina e la 'musicalità' del sistema, ovvero la possibilità di realizzare sonificazioni efficaci.

Per quanto riguarda gli sviluppi futuri si potrebbe pensare di espandere il progetto orientandolo in direzione di sistemi installativi ad uso educativo creando un applicativo Web che consenta agli utenti di interagire tramite dispositivo mobile o tablet via browser. In questo caso ci si orienterebbe a vantaggio di una programmazione funzionale alla destinazione d'uso, in Python o altro linguaggio supportato dai browser per poter realizzare anche un sistema che renda possibile le interazioni tra gli utenti, per attivare un dialogo musicale tra performer a partire dalle suggestioni che de-

rivano dagli intrecci e relazioni tra le diverse parole digitate.

## 6. BIBLIOGRAFIA

- [1] L. Meyer, *Style and Music: Theory, History, and Ideology*, University of Chicago Press, Chicago, 1989.
- [2] G. Hauptenthal, *Geschichte der Würfelmusik in Beispielen*, Universität des Saarlande, Saarbrücken, 1994.
- [3] K. Stockhausen, "The Concept of Unity in Electronic Music", *Elaine Barkin Perspectives of New Music*, Vol. 1, No. 1. (Autumn, 1962), pp. 39-48, 1962.
- [4] M. Chion, *L'arte dei suoni fissati o La Musica Concretamente*, Edizioni Interculturali, Roma, 2004.
- [5] D. Deutsch, *Psychology of Music*, Academic Press, Inc., New York, 1982.
- [6] A. Hunt e M. M. Wanderley, "Mapping performer parameters to synthesis engines", *Organised Sound* 7(2), (Cambridge University Press), pp. 97-108, 2002.
- [7] J. Scordato, "Ariadne: un sistema per l'esecuzione musicale su tablet", *Proceedings of the 19th CIM, Trieste*, November 21-24, pp. 60-65, 2012.
- [8] D. Wessel and M. Wright, "Problems and Prospects for Intimate Musical Control of Computers", in *Computer Music Journal*, 26: 3, MIT Press Cambridge, USA, 2002.
- [9] M. M. Wanderley e N. Orio, "Evaluation of Input Devices for Musical Expression: Borrowing Tools from HCI", in *Computer Music Journal*, 26:3, pp. 62-72, 2002.
- [10] S. R. Klemmer, B. Hartmann, e L. Takayama, "How Bodies Matter: Five Themes for Interaction Design", *Proceedings of the Designing Interactive Systems 2006*, University Park, PA, USA - June 26 - 28, 2006.
- [11] V. J. Manzo, *Max/MSP/Jitter for Music: A Practical Guide to Developing Interactive Music Systems for Education and More*, Oxford University Press, Oxford, pp.151-156, 2016.
- [12] A. Hunt, M. M. Wanderley e M. Paradis, "The Importance of Parameter Mapping in Electronic Instrument Design", in *Journal of New Music Research*, 32:4, pp. 429-440, 2010.
- [13] A. Hunt e R. Kirk, "Mapping Strategies for Musical Performance", *Trends in Gestural Control of Music*, M.M. Wanderley and M. Battier, eds., 2000.

# LE NUOVE SCRITTURE MUSICALI PER IL CINEMA. STUDI DI REGISTRAZIONE, MEDIA DIGITALI E PRATICHE COMPOSITIVE

Kristjan Stopar

Università degli Studi di Udine  
kristjan.stopar@uniud.it

Roberto Calabretto

Università degli Studi di Udine  
roberto.calabretto@uniud.it

Luca Cossetini

Università degli Studi di Udine  
luca.cossetini@uniud.it

## ABSTRACT

Le nuove scritture musicali cinematografiche sempre più utilizzano *software* per la creazione musicale. Tale situazione ha reso obsolete le tradizionali prassi compositive, ormai relegate ai margini della produzione. Come scrive Hans Florian Zimmer, il computer oggi è divenuto uno strumento musicale imprescindibile «and then more virtuouistic you get on that, the better you can express your ideas» [11]. Ma, nonostante un cospicuo numero di compositori utilizzi il computer in tal senso, a oggi non si dispone di uno studio scientifico approfondito sul metodo di scrittura e sul processo tecnologico alla base della creazione di una *score*. La maggior parte dei musicisti, forte di questo nuovo ambiente informatico, ha spostato il baricentro dell'atto creativo dal piano compositivo tradizionalmente inteso all'universo del *sound desing*, come confermato da innumerevoli interviste a compositori hollywoodiani, tra cui Hans Zimmer e Tom Holkenborg [12, 25]. Si rende pertanto necessario uno studio scientifico che sappia coniugare la composizione musicale con l'utilizzo dei *software* applicandola alle immagini in movimento. Il tal modo si possono creare i presupposti per un dibattito scientifico e la seguente produzione di musica per i nuovi media.

## 1. STATO DELL'ARTE

I più acclamati compositori hollywoodiani, come Max Steiner, Bernard Herrmann, Dimitrij Tiomkin, Henry Mancini, Elmer Bernstein, ed europei, come Nino Rota, Maurice Jaubert, componevano le loro partiture cinematografiche con l'ausilio del pianoforte e del tradizionale foglio pentagrammato, annotando su brogliacci cartacei le parti per i diversi strumenti e per le sezioni orchestrali. La loro prassi compositiva non differiva da quella tradizionale. Dagli anni '50 in poi, sulla scia aperta dalla sperimentazione musicale elettronica, anche autori cinematografici avevano iniziato a utilizzare i suoni sintetici; ciononostante la tecnologia elettronica non era considerata come ausilio alla composizione. È solo a partire dagli anni '80 che i sintetizzatori hanno trovato

Copyright: © 2018 Stopar et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution License 3.0 Unported](https://creativecommons.org/licenses/by/3.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

estesa e diffusa applicazione, poiché le nuove tecnologie digitali permettevano di unire sempre più la prassi compositiva tradizionale con quella tecnologica. Si è approdati così all'utilizzo di *software* di scrittura che non soltanto interagivano con l'orchestra, ma ne simulavano le sezioni. A rendere possibile questa rivoluzione fu lo sviluppo di protocolli condivisi. Nel 1981 nasceva, infatti, il protocollo MIDI, che si rivelerà essenziale per la pratica musicale sul computer. La gestione del MIDI è oggi delegata ai *sequencer*, *software* musicali più versatili rispetto ai sintetizzatori, i modulatori e i campionatori utilizzati nelle fasi pionieristiche di questa nuova prassi compositiva. La nuova generazione di musicisti dispone oggi di postazioni di lavoro standardizzate e conosciute con l'acronimo di DAW, *Digital Audio Workstation*.

Grazie a un'interfaccia grafica e al collegamento a svariate periferiche esterne di I/O si è in grado di comporre una *score* per orchestra di notevole complessità, attraverso la creazione di cosiddetti *mockup*, bozze digitali della musica<sup>1</sup>. Ciò è reso possibile dal campionamento sonoro che permette di riprodurre i timbri reali di strumenti classici. Una diversa prassi compositiva, dunque, ove è possibile ascoltare direttamente il timbro strumentale previsto dalla partitura, 'velocizzando' così la prassi produttiva: grazie alle accurate demo digitali, il regista non deve più immaginare la colonna sonora orchestrale mentre il compositore suona i temi principali al pianoforte.

La nuova prassi ha permesso altresì a molti musicisti di sperimentare formule compositive da autodidatti, senza un'educazione accademica e in alcuni casi di una conoscenza musicale adeguata per affrontare la composizione orchestrale<sup>2</sup>. Questa mancanza viene compensa-

<sup>1</sup> Per la creazione della *score* definitiva il *mockup* può essere integrato con il materiale acustico, oppure aggiungersi a esso migliorandone alcuni aspetti per mezzo di sovrapposizioni, i cosiddetti *overdub*. In altri casi, infine, si ricorre esclusivamente al *mockup* quando il *budget* ridotto non permette la registrazione dal vivo.

<sup>2</sup> Significative in tal senso le parole dell'orchestratore Gordon Goodwin: «I say that in the era where we find ourselves now, more and more composers have no practical experience with an orchestra» [15, p. 474].

ta dalla loro sensibilità all'ascolto, alla quale si aggiunge l'interesse al *sound design* e alla ricerca timbrica.

Negli ultimi anni c'è stata un'apertura verso questo nuovo metodo di lavoro anche da parte di alcuni dei più noti autori che hanno reso pubblica la propria prassi compositiva e inaugurato un dibattito teorico sulla loro scrittura<sup>3</sup>. Non solo, la musica per film inizia ad essere ampiamente studiata anche dal punto di vista storico, estetico ed etnomusicologico. Il suo sviluppo è stato ben delineato, come le dinamiche sociali che governano il network produttivo. A tal proposito vanno considerati vari studi in lingua anglosassone [1, 2, 7, 8, 9, 10] e italiana [3, 4, 5, 15, 16, 17, 18, 19, 21]. Questi scritti contemplano anche la prassi compositiva assistita al computer, ma in maniera generica ed esclusivamente teorica, senza scendere nei dettagli con esempi pratici applicati a casi esistenti. Uno studio più approfondito in tal senso è stato parzialmente realizzato nel volume di Fred Karlin e Rayburn Wright [13], in cui si fanno riferimenti alla composizione del *mockup* e ai rapporti che intercorrono tra le figure professionali lungo la catena di produzione musicale cinematografica. I software utilizzati non vengono però presi in considerazione e, di conseguenza, è necessario estendere l'orizzonte bibliografico anche alla manualistica tecnica. I volumi di Mike Collins [6] e David Sonnenschein [23], ai quali si aggiungono i manuali dei diversi *software* reperibili in rete e i siti specializzati in informatica musicale, consentono di approfondire aspetti tecnici altrimenti non presenti nei sopraccitati testi. Tra questi va sottolineato il sito [www.soundonsound.com](http://www.soundonsound.com) che affronta in modo approfondito la natura dei *sequencer*, dei *mixer* professionali e di tutte le apparecchiature presenti nelle postazioni di lavoro degli odierni compositori cinematografici. Ciononostante, anche in questi casi la definizione della prassi compositiva risulta superficiale. Si rendono necessari ulteriori studi e approfondimenti.

In questa direzione iniziano a nascere, principalmente in ambiente anglosassone, progetti didattici innovativi in seno a importanti realtà accademiche, quali il Berklee College of Music di Boston, la USC di Los Angeles e la Royal College of Music di Londra, dove tra le diverse proposte formative compare la musica per film e nuovi media e in cui si affronta anche la composizione assistita al computer. In ambito europeo si segnalano anche i corsi della Sibelius Academy di Helsinki.

## 2. OBIETTIVI

L'obiettivo del nostro progetto di ricerca all'Università degli Studi di Udine è compiere una ricognizione e un'analisi approfondita delle diverse fasi di elaborazione di una *score* creata con il computer, dall'ideazione

fino all'ultimo *mixing* del *mockup*. Sempre più compositori che lavorano con gli strumenti e le risorse offerte dalla tecnologia informatica dimostrano, infatti, una preparazione lacunosa in merito al loro utilizzo consapevole in ambito audiovisivo. È essenziale quindi avviare una riflessione scientifica che superi lo stato dell'arte precedentemente descritto, fornendo un ausilio alla composizione di musica per gli audiovisivi.

## 3. METODOLOGIA

Al fine di raggiungere gli obiettivi preposti, è necessario anzitutto compiere una ricognizione della letteratura esistente. Deve far seguito una conoscenza approfondita e l'analisi del funzionamento e delle potenzialità di un dato *software* di composizione assistita al computer utilizzato nella produzione di colonne sonore. Per questa via si può giungere alla fase centrale della ricerca: l'utilizzo consapevole dei *software* e la conseguente creazione di una demo digitale.

### 3.1 Il sequencer

Per comprendere l'odierna prassi compositiva bisogna innanzitutto considerare alcuni aspetti teorici fondamentali legati ai *software*. Viene in aiuto il fatto che, nonostante oggi siano disponibili sul mercato di diversi *sequencer* professionali, le interfacce sono genericamente standardizzate su criteri di usabilità condivisi (diverse finestre, *editing* modellato sulle prassi classiche del montaggio per tagli e giunte, la simulazione grafica dei mixer ecc.) e comportano quindi un utilizzo quasi indistinto da parte del fruitore.

Attualmente la scelta del *sequencer* è affidata alla discrezione di ciascun autore o dell'infrastruttura *freelance*, la cosiddetta *facility*, in cui lavorano diversi compositori: se alcuni utilizzano Apple Logic Pro, altri fanno riferimento a Cakewalk Sonar e Cockos Reaper, maggiormente efficienti in ambito audiovisivo. Altri ancora si avvalgono del *sequencer* Pro Tools – storicamente in vetta tra i *software* professionali utilizzati per la produzione musicale e audiovisiva – oppure di Steinberg Cubase (o il successivo Nuendo). Cubase è stato una pietra miliare, in quanto il primo a introdurre lo standard VST (*Virtual Studio Technology*), che consente di integrare *plugin* esterni di terze parti, dedicati ad esempio all'elaborazione del segnale audio o alla gestione delle *sample library*, banche dati di suoni preregistrati gestite da un campionatore (*sampler*). Tra i più utilizzati, vanno citati Native Instruments Kontakt e Vienna Symphonic Library. Comprendere la natura di questi *software* e la loro funzione agevola la predisposizione del lavoro e la creazione di *mockup*.

### 3.2 Le sample library

Nell'ambito della ricerca si sono create 21 *demo* digitali relative a diversi macrogeneri filmici con prove di montaggio audiovisivo a scopo laboratoriale. La composi-

<sup>3</sup> Tra gli esempi si consideri il volume di Rona [20], la recente *Masterclass* on-line zimmeriana *Hans Zimmer Teaches Film Scoring* [14], o ancora gli interventi di John Powell, Klaus Badelt e Sean Callery in diversi *workshop* organizzati dalla SAG/AFTRA [22].

zione è avvenuta in *sequencer* Cubase Pro 9. La maggior parte delle librerie di suoni utilizzate fa riferimento ai cataloghi Native Instruments Komplete ed Icenis Spitfire, entrambi supportati dal sampler Kontakt 5.7. A questi si aggiunge la VSL (Vienna Symphonic Library) e alcune librerie della Casa di Produzione Output di Los Angeles.

Un utilizzo consapevole delle *sample library* orchestrali, richiede un'attenzione a dettagli di natura tecnico-musicale, per raggiungere un risultato soddisfacente all'ascolto, spesso quanto più vicino possibile a una registrazione dal vivo. In questa fase infatti vengono poste le basi per il lavoro degli *editor* e del regista che a loro discrezione potranno scegliere poi se utilizzare la bozza digitale, affidarsi *in toto* alla registrazione dal vivo, o integrare quest'ultima con alcune parti del *mockup*, al fine di migliorarne la qualità per mezzo di sovrapposizioni (*overdub*).<sup>4</sup> Importante è dunque utilizzare in maniera appropriata le diverse parti orchestrali all'interno del *software*, poiché costituiranno la base per le successive fasi di registrazione e per la generazione e la stampa della partitura e delle parti da affidare agli strumentisti coinvolti nella realizzazione della colonna sonora. Per questo motivo tutte le parti orchestrali inserite nel *mockup* dovrebbero poter essere eseguite dal vivo. Questa attenzione è cruciale: è qui che la competenza organologica del compositore gioca un ruolo determinante per evitare di creare parti di fatto ineseguibili perché trascendono le effettive possibilità tecnico-esecutive dei musicisti o i limiti fisici degli strumenti acustici.

### 3.3 La prassi compositiva

L'atto compositivo avviene nella maggior parte dei casi direttamente alla tastiera MIDI, registrando di volta in volta le diverse idee con un riscontro acustico in tempo reale del timbro dello strumento scelto. La registrazione della performance, di frequente scandita da una *click track* che funge da metronomo, può essere assistita dalla quantizzazione automatica, che allinea l'attacco delle note eseguite a una griglia preposta nell'*editor*, solitamente organizzata secondo il metro. *Click track* e quantizzazione limitano però un naturale fraseggio consono a un'esecuzione dal vivo. Se l'estetica della colonna sonora lo richiede, è possibile pertanto disattivare la quantizzazione, mantenendo ugualmente la *click track* per un'esecuzione controllata.

Come nel caso dei brogliacci cartacei, il *mockup* può presentare una distribuzione delle tracce organizzata per

<sup>4</sup> Così si esprime il compositore Cristophe Beck: «So many sampled elements that are intended to be replaced by live instruments end up staying in» [13, p. 167]. Inoltre bisogna considerare che l'utilizzo del materiale digitale è obbligatorio nel caso in cui il *budget* ridotto non permetta la registrazione dal vivo. Rona scrive: «Simulating an orchestra with samples and synths is a tough business. When there is no money for a large group of players, but the project calls for an orchestral approach, then working with samples is simply the only choice» [20, p. 107].

famiglie orchestrali. Tuttavia, nel *mockup* è possibile ridurre intere sezioni orchestrali in una singola traccia di *samples*, perdendo così la divisione in parti. Ciò avviene ad esempio nel caso di uno *string ensemble* o *brass ensemble*.

Essenziale è inoltre considerare le tecniche esecutive di ogni strumento. Per quanto riguarda gli archi, ad esempio, andranno considerate le arcate e l'articolazione. Questa sezione ha una vasta possibilità timbrica, generalmente considerata dalle *library* e definibile nei diversi eventi MIDI: dai più utilizzati e comuni *staccato*, *legato*, *sforzato*, *pizzicato*, *col legno*, sino a tecniche più particolari come *bartokslap*, *wrong side of the bridge*, ecc. Si può giungere così a esecuzioni artificiali simili a quelle dal vivo. Altrettanta cura va riservata alle sezioni dei legni e degli ottoni, per le quali va considerato il respiro degli esecutori. Prevedere troppe note dai valori lunghi in legato, ad esempio, può portare a soluzioni anche musicalmente efficaci se riprodotte al *sequencer*, ma ineseguibili dal vivo. Infine, si deve porre attenzione al portamento e alle diverse fasi dell'involuppo del suono (ADSR), modificando manualmente i parametri dell'evento MIDI o sfruttando periferiche esterne quali pedali di volume e di controllo, *modulation wheel*, *pitch bender*.

Un'ulteriore difficoltà è data dal fatto che comporre eseguendo le parti alla tastiera MIDI significa trascendere dalle estensioni<sup>5</sup> e dalle tecniche esecutive dei singoli strumenti. Alcune figurazioni ritmiche, successioni accordali o arpeggiati sono opportuni per la tecnica pianistica, ma spesso molto distanti da quella di altre famiglie strumentali<sup>6</sup>. Una scrittura inconsapevole in fase di creazione del *mockup* spesso conduce alla necessità di rivedere la strumentazione in fase di stesura della partitura reale<sup>7</sup>.

Il lavoro di composizione si conclude con l'orchestrazione. Essa richiede una competenza musicale che un compositore autodidatta e formatosi attraverso la prassi del *sound design* e la scrittura assistita al *sequencer* difficilmente può avere. Anche per questo mo-

<sup>5</sup> Per alcuni autori, come ad esempio Beck [13, p. 164], ciò non costituisce un problema, anzi, può essere inteso come una scelta estetica.

<sup>6</sup> Spiega McKenzie: «Notes that get played in with your fingers sometimes are not notes that will work with an orchestra, [...] and so you can take the same gesture and put different notes there that will make the gesture happen. Strings, for example. There are strings that you can play on your fingers [triggering samples] that don't play well on string instruments. So you get the upwards sweep or the arpeggiated kind of feel, and you just use different notes» [13, p. 164].

<sup>7</sup> Ciò accade di frequente per la sezione degli ottoni. L'orchestratore Brad Dechter spiega: «You have an amazing amount of control with synth patches. You can make a French horn a brassy sound, and when I get it, I say, 'That really is going to sound better with the trombone'. Or, I might double them, if that's the appropriate thing. Or sometimes, and no offense to the composers, they'll just keep using that nice French horn patch. But, truly, it's like a low rich 1st, 5th, and 3rd trombone voicing» [13, p. 477].

tivo questa fase è demandata a un'altra figura professionale: l'orchestratore<sup>8</sup>.

#### 4. CONCLUSIONI

Le modalità di lavoro qui presentate sono state parzialmente chiarite pubblicamente, con risultati molto positivi e incoraggianti. Basti pensare che nel 2014 John Powell, tra i più attivi compositori di Hollywood, ha presentato il proprio metodo di lavorazione del *mockup* per il film *Ice Age* (2002) in un workshop organizzato dalla SAG/AFTRA, associazione che rappresenta giornalisti, attori, presentatori e altri artisti, tra i quali anche i compositori cinematografici [22]. Parimenti hanno fatto Klaus Badelt e Sean Callery nel 2013. Di recente, invece, Zimmer ha illustrato il suo metodo compositivo con *Masterclass* on-line. Nonostante la grande importanza e utilità di simili iniziative, in Europa e specialmente in Italia simili progetti non sono stati ancora promossi.

Questo progetto di ricerca può pertanto costituire un momento importante per integrare nell'universo accademico una prassi ormai consolidata e sempre più diffusa, ma mantenuta ai margini della riflessione scientifica<sup>9</sup>. In tal modo si possono creare le basi per la nascita di corsi di produzione musicale e di un nuovo ambito di ricerca, in grado di comprendere la pratica della composizione musicale per il cinema assistita al computer, come avviene in altre importanti realtà accademiche all'estero, principalmente oltreoceano. L'integrazione delle pratiche di composizione, scrittura, *editing* e *misaggio* digitali all'interno di un'istituzione accademica, permetterà di offrire corsi professionalizzanti e funzionali, per istituire innovativi centri produttivi nell'universo della musica per film.

#### 5. REFERENCES

[1] Audissino, Emilio, *John Williams's Film Music*, The University of Wisconsin Press, Wisconsin, 2014.  
 [2] Bellis, Richard, *The Emerging Film Composer: An Introduction to the People, Problems and Psychology of the Film Music Business*, BooksurgeLlc, North Charleston (SC), 2006.  
 [3] Calabretto, Roberto, *Antonioni e la musica*, Marsilio, Venezia, 2012.

<sup>8</sup> La prassi produttiva cinematografica vede un suo naturale passaggio dallo *sketch* del compositore alla rilettura dell'orchestratore, il cui lavoro risiede oggi in una trascrizione dal linguaggio MIDI alla notazione musicale tradizionale, con intrinseche difficoltà tecnico-musicali, in parte già chiarite. Il lavoro dell'orchestratore, certamente fondamentale nella catena produttiva, apre un'importante parentesi storiografica, musicologica ed etnomusicologica, qui impossibile da sviluppare. Per un approfondimento si veda [13, 15, 24].

<sup>9</sup> Al fine di rendere il progetto ancor più rilevante in ambito accademico, i risultati raggiunti saranno discussi all'interno di un convegno e sottoposti al vaglio di un'attenta analisi e riflessione per poi giungere alla redazione di un volume che costituirà un manuale per i compositori di musica cinematografica.

[4] Calabretto, R., *Lo schermo sonoro, La musica per film*, Marsilio Editori, Venezia, 2010.  
 [5] Cano, Cristina e Potì, Paolo, *Michael Nyman: ascoltare il cinema*, Franco Angeli Editore, 2004.  
 [6] Collins, Mike, *Pro Tools for Music Production: Recording, Editing, and Mixing*, Focal Press, New York, Londra, 2001.  
 [7] Cooke, Mervyn, *The Hollywood Film Music Reader*, Oxford University Press, New York, 2010.  
 [8] Davis, Richard, *Complete Guide to Film Scoring. The Art and Business of Writing Music for Movies and TV*, Berklee Press., Boston, 1999.  
 [9] Doerschuk, Robert L., *Interview with Danny Elfman*, "Keyboard", October 1989.  
 [10] Dupuis, Mauricio, *Jerry Goldsmith: Music Scoring For American Movies*, Dmg Publishing, Buckinghamshire, 2014.  
 [11] *Hans Zimmer on The Hans Zimmer Sound*, 2012-  
<http://soniccontrol.tv/2012/03/01/hans-zimmer-on-the-hans-zimmer-sound/>, ultimo accesso 20 settembre 2014.  
 [12] *IMS Engage 2014: Hans Zimmer In Conversation With Junkie XL* -  
<https://www.youtube.com/watch?v=8rrXnbsTOTQ>, ultimo accesso 16 settembre 2018.  
 [13] Karlin, Fred; Wright, Rayburn, *On the Track: a Guide to Contemporary Film Scoring*, Routledge, New York-London, 2004, pp. 164-477.  
 [14] [masterclass.com/classes/hans-zimmer-teaches-film-scoring](http://masterclass.com/classes/hans-zimmer-teaches-film-scoring), ultimo accesso 22 settembre 2018.  
 [15] Meandri, Ilario *La fabbrica dei sogni*, Kaplan, Torino, 2012.  
 [16] Miceli, Sergio e Morricone, Ennio, *Comporre per il cinema. Teoria e prassi nella musica per film*, a cura di Laura Gallenga, Marsilio, Venezia, 2001.  
 [17] Miceli, S., *Morricone, la musica, il cinema*, Ricordi, Milano, 1994.  
 [18] Miceli, S., *Musica e cinema nella cultura del Novecento*, Sansoni, Milano, 2000.  
 [19] Miceli, S., *Musica per film. Storia, estetica, analisi, tipologie*, LIM, Milano, 2009.  
 [20] Rona, Jeff, *The Reel World: Scoring for Pictures*, Hal Leonard Publishing Corporation, Milwaukee, 2001, p. 107.  
 [21] Saba, Cosetta G., *Lo sguardo che insegue*, Lupetti, Milano, 2006, pp. 28-42.  
 [22] [sagaftra.org/content/about-us](http://sagaftra.org/content/about-us), ultimo accesso 15 settembre 2015.  
 [23] Sonnenschein, David, *Sound Design: The Expressive Power of Music, Voice and Sound Effects in Cinema*, Michael Wiese Productions, Los Angeles, 2001.  
 [24] Stopar, Kristjan, *Hans Florian Zimmer. Analisi di un fenomeno hollywoodiano tra musica e cinema*, Tesi di Dottorato internazionale di studi audiovisivi: cinema, musica e comunicazione - ciclo xxviii, Università degli Studi di Udine,

Facoltà di Lettere e Filosofia, Anno Accademico  
2015-2016.

- [25] Vary, Adam B., *Inside The Mind (And Studio) of  
Hollywood's Music Maestro-*  
[http://www.buzzfeed.com/adambvary/hans-  
zimmer-film-composer-inside-his-  
studio#.icJD6y1xPG](http://www.buzzfeed.com/adambvary/hans-zimmer-film-composer-inside-his-studio#.icJD6y1xPG), ultimo accesso 20 settembre  
2018.

# MOTION CAPTURE E STRUMENTI AUMENTATI: CONTROLLO GESTUALE E GESTO COMPOSITIVO

Luca Richelli

Sound and Music Processing Lab  
Conservatorio "C. Pollini" - Padova  
lucarichelli@gmail.com

## ABSTRACT

Questo articolo presenta il lavoro di ricerca e produzione, effettuato presso il Sound and Music Processing Lab (SaMPL) del Conservatorio di Padova in collaborazione con il Centro di Sonologia Computazionale (CSC) del Dipartimento di Ingegneria dell'Informazione (DEI) dell'Università di Padova, con il sistema di *motion capture* PhaseSpace. Saranno illustrate le caratteristiche tecniche del sistema utilizzato, le produzioni musicali realizzate fino ad oggi, i problemi di ordine tecnico e musicale inerenti al lavoro svolto e le prospettive future.

L'ambito principale di utilizzo del sistema è stato l'integrazione dei parametri di rilevamento del movimento degli esecutori strumentali con l'opera musicale stessa, o visto da un punto di vista diametralmente opposto, la creazione di iper-strumenti (o strumenti aumentati) che permettono all'esecutore di controllare, oltre all'emissione acustica dello strumento tradizionale, nuove possibilità espressive elettroacustiche. Il gesto esecutivo integra la dimensione acustica ed elettronica dell'opera musicale diventando parte integrante della partitura, richiedendo al compositore di pensare anche a questa nuova dimensione.

## 1. INTRODUZIONE

Il termine *motion capture* (Mocap) indica il processo di registrazione della posizione spaziale di particolari dispositivi, denominati *marker*, da parte di un sistema hardware/software, generalmente costituito da un computer server a cui sono collegate un numero variabile di videocamere o altri dispositivi per il rilevamento dei *marker* stessi. Applicando un numero sufficiente di *marker* al corpo umano, o a qualsiasi altro corpo mobile, è possibile rilevarne, in maniera dettagliata, la posizione spaziale ma soprattutto il movimento delle singole parti che lo compongono.

Il campo di utilizzo più spettacolare e conosciuto è l'intrattenimento per la creazione, tramite animazione computerizzata, di effetti speciali in film e videogiochi,

*Copyright: © 2018 Luca Richelli. This is an open-access article distributed under the terms of the Creative Commons Attribution License 3.0 Unported, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.*

ma il *motion capture* è presente in ambito militare, medico, biomeccanico, sportivo e di studio e ricerca.

I primi sistemi di rilevamento del movimento risalgono agli anni settanta come strumento di analisi fotogrammetrica nella ricerca biomeccanica; nel corso degli anni sono state impiegate tecnologie acustiche, meccaniche, magnetiche e ottiche [9, 15] anche senza *marker* (*markerless*) [14], e sistemi ibridi. Inizialmente l'elaborazione dei dati avveniva in tempo differito e quindi utilizzabile principalmente per l'analisi e la post-produzione.

Si deve aspettare l'inizio del nuovo millennio per l'elaborazione in tempo reale dei dati, e viene comunemente indicata la trasposizione cinematografica del *Signore degli Anelli*<sup>1</sup> (2001-2003) come la prima realizzazione efficace e convincente di *motion capture*. Successivamente si è assistito ad un diffondersi sempre più massiccio di strumenti di rilevazione del movimento, soprattutto in ambito ludico, per l'interazione con la realtà virtuale.

## 2. PHASESPACE

Poco dopo la sua costituzione – 2010 – il SaMPL<sup>2</sup> si è dotato di uno dei sistemi di rilevamento del movimento più all'avanguardia: il PhaseSpace<sup>3</sup>, un sistema di *motion capture* ottico a LED attivi. La configurazione installata presso la sede del Centro di Sonologia Computazionale (CSC) del Dipartimento di Ingegneria dell'Informazione (Università di Padova) è costituita da un server a cui sono collegate 8 videocamere all'infrarosso da 3600x3600 pixel – 12 Megapixel - che utilizzano due sensori lineari ad alta risoluzione a 16 bit con una frequenza di 480 fotogrammi al secondo ed una latenza inferiore ai 10 millisecondi. Il vantaggio principale dei LED attivi<sup>4</sup> – figura 1 – è di essere riconosciuti univocamente dal sistema grazie ad una diffe-

<sup>1</sup> <https://www.theguardian.com/film/2002/nov/07/lordoftherings.seanclarke>

<sup>2</sup> <http://www.sampl-lab.org>

<sup>3</sup> <http://phasespace.com/>

<sup>4</sup> Questa caratteristica dei sistemi a LED attivi li differenzia in modo sostanziale dai sistemi software di video tracking - EyesWeb, Jitter, ... - che riducono/analizzano a due dimensioni la scena reale - e che spesso confondono immagini tra loro speculari come ad esempio la sagoma frontale e posteriore di una persona.

rente modulazione del segnale. I LED vengono collegati ad un LED controller tascabile che trasmette i dati in radio frequenza (RF) al ricevitore - *LED Base Station* - collegato al server.



**Figure 1.** LED attivo. Dimensioni 20 x 14 x 3.2 mm

Il sistema rileva ogni LED con una terna di coordinate cartesiane  $[x, y, z]$  ed applicando un numero sufficiente di sensori è possibile ricostruire con notevole precisione i movimenti degli arti del corpo umano: utilizzo tipico per analisi biomeccanica o per creazione, in post-produzione, di avatar per film e videogiochi. Nonostante l'affidabilità del sistema, che nella configurazione installata, ha a disposizione otto videocamere per individuare i LED, è possibile che alcuni marker siano momentaneamente non rilevabili a causa del mascheramento da parte del performer stesso o dello strumento musicale utilizzato. Considerando che, a differenza degli impieghi di analisi e/o post-produzione, dove è possibile ripetere più volte la sequenza dei movimenti, in uso performativo del *motion capture* – concerto – non è pensabile fermare e ripete l'esecuzione. Di conseguenza, per aumentare l'affidabilità del tracciamento, si è deciso di creare dei *rigid body*: oggetti fisico-virtuali<sup>5</sup> composti da più marker, ma tracciabili come unità. Nello specifico sono stati messi a punto due tipologie di *rigid body*: *glove* (guanto mezza dita<sup>6</sup> a 4 LED) e *bangle* (bracciale a 7 LED). La condizione minima di tracciabilità di entrambi è l'individuazione di almeno un LED, inoltre è possibile rilevare un ulteriore parametro: la rotazione (*R*).

### 3. PRODUZIONI MUSICALI

#### 3.1 Musica Elettronica e Gesto

Il rapporto tra musica elettronica e gesto è particolarmente complesso ed articolato. In relazione al periodo storico, ambito stilistico, contesto tecnologico e forma/mezzo utilizzati, semplificando e schematizzando,

<sup>5</sup> Il rigid body è costituito da un oggetto fisico - generalmente un indumento: guanto, bracciale, cappello, scarpa, tuta, ... - a cui vengono attaccati dei LED in posizioni stabilite. Conoscendo la disposizione dei LED sul rigid body fisico - in particolare la distanza fra di loro - è possibile tracciare il movimento della controparte virtuale anche in mancanza di alcuni LED. Ovviamente maggiore è il numero dei LED maggiore è l'affidabilità del sistema.

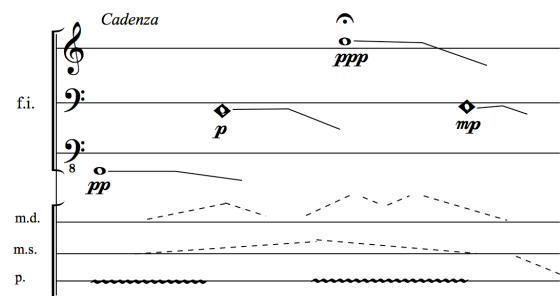
<sup>6</sup> Lasciando libera la parte terminale delle dita, il guanto a mezza dita può essere impiegato da differenti strumentisti senza ostacolare eccessivamente la tecnica acquisita.

il gesto può essere: a) assente nelle realizzazioni per *fixed media*, b) retroattivo, paritario al gesto strumentale tradizionale, quando i parametri fisici sono direttamente collegati al risultato sonoro, c) metacontestuale, con o senza *feedback*, quando la relazione di causa ed effetto non è esplicita ma demandata ad una sovrastruttura di controllo. In questa prospettiva i sistemi di rilevamento del movimento possono essere considerati come un'evoluzione delle periferiche di controllo uomo-macchina [6].

La ricerca, intrapresa al SaMPL sull'utilizzo di questa nuova tecnologia, si è subito collegata alla produzione musicale, cercando di avere un quadro più ampio possibile di molteplici approcci compositivi [1, 2]. La collaborazione con differenti compositori e performer ha portato alla realizzazione delle opere di seguito esaminate.

#### 3.2 Roberto Fabbriciani *Motion capture*

In *Motion capture*<sup>7</sup> di Roberto Fabbriciani per flauto iperbasso<sup>8</sup>, live electronic e mocap [11] il compositore-esecutore ha notato in partitura, oltre alla parte strumentale, anche la parte gestuale di controllo delle elaborazioni elettroacustiche. L'opera nasce proprio dalla sperimentazione dell'esecutore sull'allargamento delle possibilità timbriche ed espressive dello strumento. Il flauto iperbasso non possiede chiavi<sup>9</sup>, l'esecutore ha le mani libere, non vincolate da necessità esecutive, ed è stato possibile realizzare un ambiente esecutivo elettroacustico partendo proprio dai dati rilevabili dal PhaseSpace



**Figure 2.** Roberto Fabbriciani *Motion capture*<sup>10</sup>. Estratto della partitura

<sup>7</sup> Prima esecuzione assoluta 20 maggio 2010 Teatro Sociale, Cittadella (PD). È possibile vedere un breve estratto dell'opera a <http://www.youtube.com/watch?v=Bty8KVf0Js8>

<sup>8</sup> Costruito nel 2000 da Roberto Fabbriciani, il flauto iperbasso è il più grande strumento della famiglia dei flauti. La sua lunghezza è di 15.30 metri e la nota più bassa che può emettere è  $do^{-1}$  -  $C^0$  in notazione americana standard. Per maggiori approfondimenti [https://it.wikipedia.org/wiki/Flauto\\_iperbasso](https://it.wikipedia.org/wiki/Flauto_iperbasso)

<sup>9</sup> Solo alcune valvole di apertura/chiusura con funzione e utilizzo più simili ai registri organistici che alle chiavi flautistiche. Permettono di cambiare la serie armonica delle note fondamentali:  $do^{-1}$ ,  $sol^0$ ,  $re^1$ ,  $la^1$ ,  $mi^2$ ,  $re^1$ ,  $do^3$ . Le variazioni timbriche derivano principalmente dal controllo ed articolazione dell'immissione dell'aria nello strumento.

<sup>10</sup> Copyright Edizioni Suvini Zerboni - Sugarmusic, Milano. Per gentile concessione dell'editore.

Tramite due *rigid body glove*, indossati dall'esecutore, viene tracciato il movimento delle due mani associati al controllo d'altezza, del timbro e della proiezione dei suoni nello spazio. Nello specifico la mano destra controlla: asse X (destra-sinistra) trasposizione del suono nell'ambito di due ottave, asse Y (alto-basso) brillantezza timbrica tramite filtro shelf, asse Z (avanti- dietro) inarmonicità del suono; la mano sinistra: asse X (destra-sinistra): spazializzazione destra sinistra, asse Z (avanti-dietro) spazializzazione posteriore-frontale. Oltre ai controlli gestuali l'esecutore ha a disposizione un pedale per l'attivazione del banco dei delay.

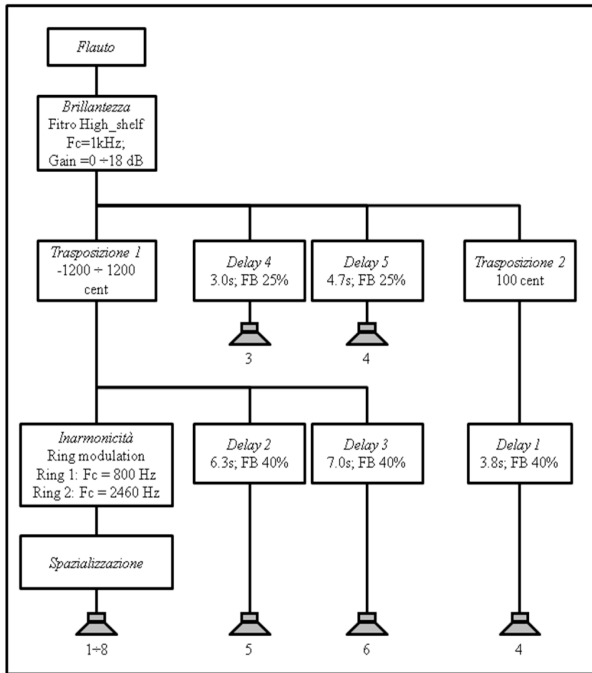


Figure 3. Schema a blocchi dell'ambiente esecutivo

L'utilizzo del PhaseSpace in *Motion capture* è altamente integrato sia con la scrittura musicale che con il gesto esecutivo. Da un punto di vista compositivo le trasformazioni elettroacustiche, e il loro controllo gestuale, sono parte integrante della partitura, la cui stesura nasce proprio dalla composizione dei movimenti delle mani. Da un punto di vista tecnico il tracciamento è particolarmente affidabile e permette all'esecutore una grande libertà di interpretazione.

### 3.3 Claudio Ambrosini *Ogni Emozione dell'Aria*

In *Ogni emozione dell'aria*<sup>11</sup>, per clarinetto basso, live electronics e mocap, [10] il compositore, Claudio Ambrosini, ha condizionato la scrittura strumentale al funzionamento dei *rigid body bangle*. Nei momenti in cui è richiesto il controllo delle trasformazioni elettroacustiche da parte dell'esecutore, la scrittura strumen-

<sup>11</sup> Prima esecuzione assoluta 7 luglio 2011, SMC2011, Auditorium Pollini Padova. L'intera esecuzione è visibile a <https://www.youtube.com/watch?v=-1M19rPnbiY&feature=youtu.be>

tale utilizza solamente note che si possono emettere sul clarinetto basso senza l'utilizzo delle mani o con una mano sola: destra o sinistra. In questo modo le mani possono eseguire i movimenti indicati in partitura.

L'ambiente esecutivo informatico, realizzato con Max da Alvisè Vidolin, attua le indicazioni del compositore mappando i movimenti delle mani del clarinetista utilizzando le tre coordinate spaziali [x, y, z] e il valore modulo [M] delle coordinate stesse<sup>12</sup>. Nella composizione sono previsti tre differenti configurazioni di controllo (CC1-3): solo mano destra (RH), solo mano sinistra (LH), entrambi le mani (RLH).

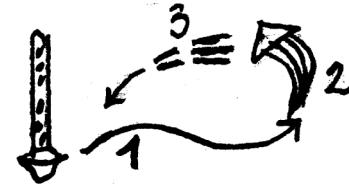


Figure 4. Claudio Ambrosini *Ogni emozione dell'aria* Annotazione in partitura: “(Mano dx) Di nuovo fa uscire il suono dal clar., lo fa alzare di altezza e lo rimette dentro

Le tre configurazioni e la corrispondenza tra i movimenti delle mani con le trasformazioni elettroacustiche sono riportati nella tabella seguente.

| Conf. | Corrispondenza movimento-elaborazione   |
|-------|---|
| CC1   | XR: Dissolution_A output level <sup>13</sup><br>YR: Dissolution_A trasposition<br>ZR: space <sup>14</sup> Left-Right<br>MR: space Front-Rear  |
| CC2   | XL: Dissolution_A output level<br>YL: Dissolution_A trasposition;<br>ZL: space Left-Right<br>ML: space Front-Rear   |
| CC3   | YR: Dissolution input level; Dissolution_B level <sup>15</sup><br>ZR: Dissolution_B reverberation level; Ring level<br>XL: Dissolution_A output level<br>YL: Dissolution_A trasposition; space Left-Right<br>ZL: space Front-Rear |

Table 1. Mappatura Phasespace – Max

Ulteriore conseguenza di questa scelta tecnica e compositiva è la valenza coreografica che l'esecuzione strumentale acquisisce. I gesti delle mani del clarinetista accompagnano il suono e le sue trasformazioni, in

<sup>12</sup> Il modulo è uguale alla radice quadrata della somma dei quadrati delle singole coordinate.

<sup>13</sup> Filtraggio FFT a soglia, con modifica dell'involuppo d'ampiezza e trasposizione, modulazione ad anello del segnale originale, equalizzazione (highshelf) della somma dei segnali.

<sup>14</sup> Controllo dello Spazializzatore ottonico

<sup>15</sup> Granulazione e trasposizione mediante FFT

una danza gestuale che esplicita le indicazioni testuali del compositore: far uscire il suono dallo strumento, alzarlo in altezza e mandarlo verso gli ascoltatori.

### 3.4 Nicola Sani *No Landscape*

Nicola Sani in *No Landscape*<sup>16</sup>, per pianoforte, live electronics e mocap, ripristina il ruolo centrale dell'interprete strumentale nella produzione elettroacustica. Con l'avvento del live electronics l'interpretazione dello strumentista è stata generalmente limitata alla componente acustica, affidando la controparte elettronica alla regia del suono. Grazie alla tecnologia di *motion capture* lo strumentista può ora integrare la componente elettroacustica con quella acustica riunificando il gesto musicale.

In questo caso i parametri ricavati da PhaseSpace derivano, quasi completamente, dall'esecuzione strumentale: il pianista, indossa il *rigid body glove* ed esegue la partitura in modo tradizionale senza avere ulteriori indicazioni di movimento. È dal movimento pianistico stesso che vengono scelti dal compositore alcuni parametri spaziali per le trasformazioni elettroniche. Le sfumature agogiche e dinamiche, che differenziano le esecuzioni tra loro, si riflettono, perfettamente sincronizzate, anche nell'esecuzione elettroacustica. Ogni interprete caratterizzerà in modo personale ed unico, attraverso la propria getualità, anche le trasformazioni elettrocustiche.

Nel corso della composizione i parametri spaziali sono mappati come da tabella seguente:

| Coordinate | Parametri di controllo  |
|------------|---|
| XL         | ambito di trasposizione in FFT<br>tempo di ritardo del banco di delay   |
| YR         | frequenza filtro passa banda,<br>spazializzatore 1-2 (destra sinistra),<br>spazializzatore 5-6 (destra sinistra)  |
| ZR         | guadagno del filtro high shelf<br>spazializzatore 1-2 (avanti indietro),<br>spazializzatore 5-6 (avanti indietro) |

**Table 2.** Mappatura Phasespace - Max

### 3.5 Luca Richelli *PianoSpace*

L'installazione *PianoSpace*<sup>17</sup> [13], realizzata dall'autore del presente articolo, esplora le potenzialità sonore del pianoforte attraverso la creazione di molteplici Modelli Morfologici Virtuali<sup>18</sup> e differenti Mappe Composi-

<sup>16</sup> Prima esecuzione assoluta 21 ottobre 2012 "Ex Novo Musica Stagione 2012", Gran Teatro La Fenice, Sale Apollinee - Venezia.

<sup>17</sup> Presentata il 27 settembre 2013 a "Le Visioni del suono del Centro di Sonologia Computazionale" - Dipartimento di Ingegneria dell'Informazione - Notte dei Ricercatori Palazzo Bo, Padova. Un breve estratto dell'installazione è visibile a <https://www.youtube.com/watch?v=1NCc7spSU14&t=6s>

<sup>18</sup> Per Modello Morfologico Virtuale, l'autore intende il superamento del paradigma esecutivo tradizionale in cui i limiti fisici delle mani

tive<sup>19</sup>. Il sistema è formato, oltre al PhaseSpace, da un Disklavier<sup>20</sup>, da un computer con un software per la gestione dati (Max) e da un'interfaccia MIDI.

L'implementazione di un Modello Morfologico Virtuale, che si basa sui dati rilevati dai *rigid body glove*, ha permesso di rovesciare il paradigma esecutivo classico, in cui è movimento delle mani che si deve adattare alle caratteristiche fisico-morfologiche dello strumento musicale: in *PianoSpace* è lo strumento che si adatta al movimento dell'esecutore.

L'esecutore indossa due *rigid body glove* e si muove liberamente nello spazio delimitato dalle videocamere. Il flusso dei dati di rilevamento spaziale segue il percorso: PhaseSpace - protocollo OSC - computer/Max - Protocollo MIDI<sup>21</sup> - Disklavier.

L'atto compositivo consiste nella realizzazione delle Mappe Compositive che trasformano i dati del Modello Morfologico Virtuale nei parametri esecutivi per il Disklavier ri assemblando la disposizione delle note della tastiera all'idea compositiva. In altri termini non è più l'idea compositiva che si adatta ai limiti fisici dell'esecutore vincolato dalle caratteristiche morfologiche dello strumento ma lo strumento che si modifica in funzione dell'idea compositiva.<sup>22</sup>

La tabella seguente riporta la corrispondenza tra i movimenti nello spazio dei sensori e i relativi parametri del Modello Morfologico Virtuale.

| Movimento spaziale   | Parametro controllato nel Modello Morfologico Virtuale |   |
|----------------------|--|---|
|                      | <i>glove</i> destro                                    | <i>glove</i> sinistro                   |
| X (Sinistra- Destra) | Pivot Note   | Keyboard Configuration                  |
| Y (Alto- Basso)      | Note on/off  | Damper Pedal on/off                     |
| Z (Avanti- Indietro) | Pattern Chord  | Key Velocity                            |
| R (Rotazione)        | Density  | Pattern Chord Key Velocity Distribution |

**Table 3.** Mappatura Phasespace – Disklavier

delle mani dell'esecutore - estensione, velocità, polifonia, grado di differenziazione dinamica, ... - condizionano la scrittura musicale. Tramite il Modello Morfologico Virtuale, il controllo dello strumento è svincolato dalle caratteristiche fisico-morfologiche della tastiera.

<sup>19</sup> Per Mappa Compositiva si intende la trasformazione dei dati del Modello Morfologico Virtuale in parametri esecutivi per il Disklavier.

<sup>20</sup> Pianoforte ibrido prodotto dalla Yamaha che usa dei solenoidi elettro-meccanici e dei sensori ottici per un controllo meccanizzato della martelliera del pianoforte.

<sup>21</sup> Parametri esecutivi relativi a: note on/off, key velocity e damper pedal on/off.

<sup>22</sup> Per esempio l'estensione massima di una mano è di circa una decima, ma ri assemblando il pianoforte posso disporre i tasti in modo di poter eseguire con una mano qualsiasi intervallo; questo è quello che fanno le Mappe virtuali: adattano lo strumento in funzione della scrittura musicale.

### 3.6 Analisi delle opere

I casi di studio presentati illustrano quattro differenti visioni dell'utilizzo del *motion capture* sia da un punto di vista esecutivo che compositivo.

In *Motion capture* la scelta del flauto iperbasso consente al compositore di disporre completamente del movimento delle mani e di conseguenza la mancanza di vincoli strumentali permette una scrittura gestuale molto articolata ed integrata con l'esecuzione: non sarebbe possibile eseguire il brano senza un sistema di *motion capture* ma non necessariamente PhaseSpace.

In *Ogni Emozione dell'Aria* la scrittura strumentale condiziona pesantemente la scrittura gestuale: sono poche le note che si possono emettere con un clarinetto basso senza la mano destra, sinistra o entrambi, qualsiasi altra composizione avrebbe le stesse caratteristiche: solo alcune note permettono il *motion capture* mentre tutte le altre ne sono escluse.

L'esecuzione pianistica, rispetto a quella clarinetistica, permette naturalmente di avere una serie di parametri spaziali delle mani correlati alla partitura: le mani si spostano sulla tastiera seguendo la disposizione dei tasti da sinistra verso destra, dal grave all'acuto, e proprio per questo motivo in *No Landscape* non sono indicati in partitura movimenti particolari: le mani si spostano necessariamente lungo la tastiera. Anche nelle sezioni in cui il pianista suona direttamente sulla cordiera<sup>23</sup> il movimento delle mani è correlato al gesto strumentale. Le trasformazioni elettroacustiche enfatizzano proprio il gesto naturale del pianista. In questa composizione il *motion capture* raggiunge un alto grado di integrazione, e sebbene *No Landscape* sia stato recentemente eseguito<sup>24</sup> con un sistema di *motion capture* video<sup>25</sup>, l'utilizzo del PhaseSpace si è confermato più solido e affidabile.

Da un punto di vista compositivo *PianoSpace*, per il suo carattere installativo, si differenzia in modo sostanziale dalle opere precedenti. Risulta inoltre difficile stabilire un confine preciso tra la parte compositiva, esecutiva e di liuteria elettronica, non solo perché l'autore è sempre la stessa persona, ma soprattutto per l'integrazione completa tra le varie parti: il sistema può eseguire solamente la propria specifica partitura che è anche il proprio sistema di *motion capture* e di trasferimento dati. In questo caso il controllo tramite *motion capture* è indispensabile e difficilmente realizzabile senza un sistema a LED attivi come il PhaseSpace.

## 4. SVILUPPI FUTURI

La ricerca sviluppata fin'ora al SaMPL è stata orientata principalmente alla realizzazioni di iper-strumenti e alla

<sup>23</sup> La partitura prevede in alcune sezioni di percuotere le corde e/o l'interno del pianoforte con oggetti differenti e di strofinare con della pece dei fili di nylon predisposti attorno a determinate corde per ottenere un effetto di corda suonata con l'arco.

<sup>24</sup> 25 luglio 2018, Chiesa di San Agostino, Chigiana International Festival & Summer Academy, Siena

<sup>25</sup> Versione di motion tracking realizzata con MAX/Jitter

produzione musicale. Altri ambiti di ricerca con grandi possibilità di sviluppo sono l'analisi del gesto musicale e la sonificazione gestuale.

L'impiego del *motion capture* nell'analisi del gesto musicale può integrarsi con i modelli esistenti – acustico, ottico, statistico, percettivo, semantico ... [3, 5, 7] – per la creazione di una banca dati multimediale contenente file audio, video e di tracciamento gestuale del repertorio musicale. I dati raccolti potranno servire sia per lo sviluppo di modelli specifici che da punto di riferimento per l'ulteriore sviluppo di iper-strumenti. Ad oggi le esperienze compositive realizzate non hanno ancora creato una base di metodologie condivise tra i compositori; la stesura di un protocollo di analisi gestuale tramite *motion capture* potrà portare ad una formalizzazione dell'analisi stessa.

La sonificazione gestuale [8] è forse il campo di ricerca più ambizioso: per le infinite direzioni che può intraprendere, per le competenze trasversali che richiede<sup>26</sup> e per le risorse finanziarie necessarie. L'integrazione dei principi della *Parameter Mapping Sonification* (PMSon) [12] nel tracciamento gestuale tramite *motion capture* aiuterà a sviluppare ulteriormente la mappatura dei parametri fisici nella loro trasformazione sonora.

## 5. CONCLUSIONI

Le opere analizzate rappresentano una panoramica non esaustiva<sup>27</sup> ma molto variegata su differenti approcci compositivi dell'utilizzo del *motion capture* in ambito musicale. Come spesso accade, l'introduzione di nuove tecnologie richiede tempo per l'apprendimento sia da parte degli interpreti che dei compositori. Se da un punto di vista tecnologico dieci anni rappresentano un lasso di tempo enorme, da un punto di vista compositivo rappresentano poco più di un istante.

Nel frattempo nel campo del tracciamento del movimento in tempo reale sono stati commercializzate sia interfacce hardware – Microsoft Kinect, Nintendo Wiimote, Sony Move, Leap Motion – che piattaforme software – Eyesweb [4], Max/Jitter – che consentono il riconoscimento del movimento con costi irrisori rispetto ai sistemi professionali quali il PhaseSpace. Le prestazioni di questi sistemi a basso costo non raggiungono ovviamente la precisione, il dettaglio e l'affidabilità dei sistemi professionali, ma ne consentono comunque un utilizzo in ambito performativo, quando il numero di esecutori e di parametri richiesti è relativamente basso.

<sup>26</sup> L'elenco delle categorie che possono essere coinvolte è praticamente infinito. Oltre a compositori e tecnici informatici, chiunque abbia a che fare con il movimento può contribuire: attori, ballerini, coreografi, mimi, sportivi, ...

<sup>27</sup> Altre opere che hanno utilizzato il PhaseSpace del SaMPL sono state: A. Guarnieri, *...fili bianco velati...* per violino, elettronica e mocap (2010), N. Sani *More is Different* per flauto iperbasso, elettronica e mocap (2014), G. Albin *Terza Algebra del Tempo* per percussioni, elettronica e mocap (2014) oltre ad impieghi didattici, dimostrativi, seminariali e di ricerca.

Oltre al costo elevato un altro svantaggio del PhaseSpace è il tempo necessario all'installazione e taratura del sistema, che richiede, a seconda della configurazione, dalle due alle tre ore; considerando che il tempo disponibile per il montaggio e le prova prima del concerto si è ridotto in maniera significativa negli ultimi anni, questo fattore risulta cruciale per un'utilizzo *on the road* del sistema.

Il PhaseSpace è un sistema potentissimo che, per essere sfruttato in modo consono alle sue potenzialità necessita di uno spazio esclusivo dedicato sia alla ricerca che alla produzione sul modello dell'*Espace de projection* dell'IRCAM<sup>28</sup>, di personale specializzato per la gestione e di fondi per la ricerca e la produzione. Di fatto deve essere considerato alla stregua di qualsiasi altro strumento musicale: gli strumentisti impiegano anni per acquisire la tecnica necessaria per padroneggiare l'emissione sonora dei loro strumenti, i compositori impiegano anni per acquisire le conoscenze e competenze indispensabili per la scrittura musicale e così sarà anche per il *motion capture*.

Le risorse finanziarie disponibili in altri ambiti, pensiamo al mondo cinematografico, non sono minimamente paragonabili al quelle a disposizione della musica contemporanea, e questo si riflette sul livello quantitativo e soprattutto qualitativo delle opere prodotte.

## 6. BIBLIOGRAFIA

- [1] F. Bevilacqua, L. Naugle and I. Valverde, "Virtual dance and music environment using motion capture" *Proceedings of the IEEE - Multimedia Technology And Applications Conference*, Irvine CA, 2001
- [2] F. Bevilacqua, J. Ridenour and D. J. Cuccia: "3D motion capture data: motion analysis and mapping to music", *Proceedings of the Workshop on Sensing and Input for Media-centric Systems*, Santa Barbara, CA, 2002
- [3] A. Camurri, G. De Poli, A. Friberg, M. Leman And G. Volpe: "The MEGA Project: Analysis and Synthesis of Multisensory Expressive Gesture in Performing Art Applications", *Journal of New Music Research* vol. 34, pp. 5-21, 2005
- [4] A. Camurri, S. Hashimoto, M. Ricchetti, A. Ricci, K. Suzuki, R. Trocca, and G. Volpe "EyesWeb: Toward Gesture and Affect Recognition in Interactive Dance and Music Systems", *Computer Music Journal*, Vol. 24 No. 1, pp. 57-69, 2000
- [5] S. Canazza, G. De Poli, L. Mion, A. Rodà, A. Vidolin and P. Zanon "Expressive Classifiers at CSC: An Overview of the main Research Streams" *Proceedings of the XIV Colloquium on Musical Informatics*
- [6] S. Canazza, G. De Poli and A. Rodà: "CaRo 2.0: An Interactive System for Expressive Music Rendering" *Advances in Human-Computer Interaction*, Volume 2015, Article ID 850474
- [7] G. De Poli, G. D'Incà and L. Mion: "Computational Models for Audio Expressive Communication" *AES Italian Session - Annual Meeting*, Paper 05001, 2005
- [8] J. Eacott and T. Flood: "Sonification as musical performance" *Proceedings of the International Computer Music Conference*, pp. 69-74, 2011
- [9] M. Field, Z. Pan, D Stirling and F. Naghdy: "Human motion capture sensors and analysis in robotics", *Industrial Robot: An International Journal*, Vol. 38 Iss 2 pp. 163 – 171, 2011
- [10] A. de Götzen, N. Bernardini and A. Vidolin: "Musical perspectives on composition, sonification and performance." *Proceedings - 40th International Computer Music Conference and 11th Sound and Music Computing Conference, SMC 2014, ICMC 2014*, pp. 371- 378, 2014
- [11] A. de Götzen, A. Vidolin, and N. Bernardini: "Augmenting traditional instruments with a motion capture system". *Proceedings of CHI 2013 Conference on Human Factors in Computing*, 2013
- [12] T. Hermann, A. Hunt and J. g. Neuhoff, *The Sonification Handbook*, Logos Verlag, Berlin, 2011
- [13] L. Richelli, S.Canazza: "PianoSpace: Modelli Morfologici Virtuali e Mappe Compositive" *Proceedings of XX CIM*, pp. 161-163, 2014
- [14] B. Rosenhahn, C. Schmaltz, T. Brox, J. Weickert, D. Cremers, and H.-P Seidel: "Markerless motion capture of man-machine interaction", *IEEE Conference on Computer Vision and Pattern Recognition, (CVPR 2008)*, Anchorage, AK, 23-28 June, pp. 1-8, 2008
- [15] G. Welch, and E. Foxlin: "Motion tracking: no silver bullet, but a respectable arsenal", *Computer Graphics and Applications*, Vol. 22 No. 6, pp. 24-38, 2002

<sup>28</sup> <http://web4.ircam.fr/1039.html?&L=0>

# DEVELOPMENT OF REAL-TIME AUDIO APPLICATIONS USING PYTHON

**Yuri De Pra**

University of Udine  
 Departement of Mathematics,  
 Computer Science and Physics  
 Via delle Scienze 206  
 Udine 33100, Italy  
 depra.yuri@spes.uniud.it

**Federico Fontana**

University of Udine  
 Departement of Mathematics,  
 Computer Science and Physics  
 Via delle Scienze 206  
 Udine 33100, Italy  
 federico.fontana@uniud.it

**Michele Simonato**

Electrolux Professional S.p.A.  
 Viale Treviso, 40  
 Pordenone 33170, Italy  
 michele.simonato@electrolux.it

## ABSTRACT

The development of real-time audio applications for general purpose OS like Linux or MS Windows is a well-known and not trivial problem. Several real-time audio software with libraries for multi-purposes programming languages like C, C++ or Java have been developed. Although these solutions provide portability through multiple OS, they often introduce problems related to dependencies and versions alignment over the time. In the present article, the process to build a framework for real-time audio relying on the programming language Python is described. We selected Python because it offers a fast learning rate coupled with several modules that can be used inside real-time audio projects. The advantages of this programming language but also the limitations in terms of execution time imposed by the interpreted code are both discussed. In particular, a procedure of code refactoring to obtain the real-time execution is presented. This procedure involves the inclusion of compiled code generated by Cython and C language, only where strictly needed. In the last part, some case studies related to the implementation of Virtual Analog models are presented introducing different amounts of compiled code and measuring the execution time.

## 1. INTRODUCTION

In past years the programming language Python has spread over, and this is confirmed by its growth in the teaching activities at Universities [1] [2], where it is used as a powerful teaching-by-examples tool [3]. As a consequence, the usage of Python is increasing also in the developer community [4]. The community maintains hundreds of modules and libraries including tools for scientific computing (e.g. Numpy, Scipy, Matplotlib). In addition Python is the language used to develop most of the Artificial Intelligence libraries that nowadays are increasingly adopted as novel approaches in many computer music applications. The strengths of Python are the fast learning rate, the short

prototyping time and the readability of the code. By contrast, Python historically has never been considered a good candidate for real-time programming due to its interpreter that is slow in the execution of the low-level computational code. However, the usage of Cython makes it possible to embed numerical loops, running at C speed, directly in Python code [5]. Cython is an optimizing static compiler that allows to write C extensions for Python and tunes Python code into plain C performance by adding static type declarations. Thus Cython allows to translate existing algorithms designed for Matlab or C and execute them in real-time. Eventually it's needed to write C code only for the high CPU-bound code parts (e.g. arithmetical operations) minimizing the dependencies with C libraries. The aim of this work was to build a lightweight and stable framework for real-time audio minimizing external dependencies.

## 2. RELATED WORK

The usage of Python in digital signal processing (DSP) is not novel in open source projects. For instance, in 'RT-Graph' multiple physiological signals collected by sensors are displayed and real-time processed using Python and the Qt framework [6]. Concerning audio analysis tools and music information retrieval, some solutions are already available and ready to use [7] [8]. Real-time audio software based on Python already exists as well. The Pyo module [9] gives the user an easy way to build audio effect chains based on well-constructed algorithms through an object-oriented approach. Even if this solution seems easy to use, the design of custom algorithms is not trivial and the way in which the modules are built tends to hide how the data at low-level are managed. By contrast, our approach aims to give the user the possibility to deal directly with samples. We based our project only on the library PyAudio [10] that provides the bindings for PortAudio, a multi-platform library to manage the audio I/O. PyAudio provides some functions to handle real-time audio streams: this allows to work at low-level on audio samples that are grouped into chunks. The dimension of the chunks can be selected by the user to find the best compromise between the tolerable latency and the available computational resources.

*Copyright: ©2018 Yuri De Pra et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution License 3.0 Unported](https://creativecommons.org/licenses/by/3.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.*

### 3. FRAMEWORK

We approached the problem by maximizing the amount of interpreted code, to find out the limit of the solution. To identify critical or inefficient code sections, we measured the execution time of each line of the code and the number of repetitions needed to process each sound sample. These time measurements were carried out using the Python module `line profiler` [11]. This module allows the user to choose the level of the analysis (e.g. single line, function blocks) and the amount of repetitions to perform statistical analysis on the execution time of the code. We tried several buffer dimensions between 128 and 2048 samples to minimize the latency introduced to comply with the real-time execution on different architectures. The implementations have been tested using two laptop PC with AMD-A6 and Intel-i5 CPUs running Ubuntu Linux as OS. We also set different configurations of the input channel: embedded microphone, external sources and wave audio files.

#### 3.1 First steps

First of all, we developed some common functions in Python to test the functionalities of the PyAudio library and the limits of the programming language. In the following code example, we report the main structure of the audio callback required to process a chunk.

```
def callback(i_d, f_count, t_info, f):
    data = wavefile.readframes(f_count)
    samples = pcm2float(data)
    y = process(data)
    out = float2pcm(y)
    return (out, pyaudio.paContinue)
```

The functions `pcm2float(x)` and `float2pcm(x)` simply convert the format of the samples between bytes and normalized  $[-1., 1.]$  floats values. The general function `process(data)` is used to encapsulate the DSP algorithms.

After the creation of this structure, we started to build a simple low-pass filter designed as:

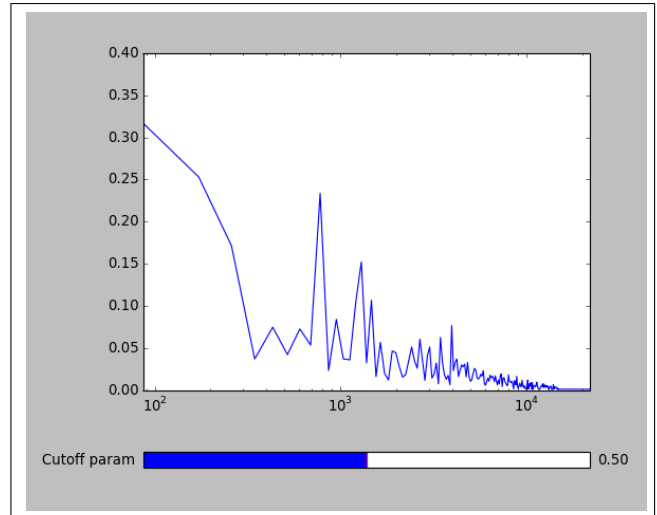
$$y[n] = \alpha x[n] + (1 - \alpha)y[n - 1] \quad (1)$$

In the following code example, we can see the simplicity of this filter implementation. We want to highlight that Python imposes to declare all the global variables that are used inside the function.

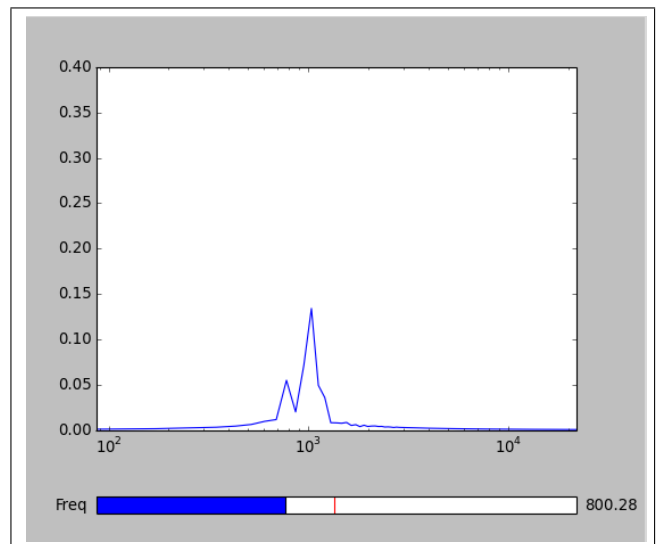
```
def process(data):
    global last_sample, a

    b = 1-a
    n= CHUNK_SIZE
    y= np.arange(n, dtype=float)

    y[0] = a*data[0] + b*last_sample
    for i in range(1,n):
        y[i] = a*data[i] + b*y[i-1]
    last_sample = y[i]
    return y
```



**Figure 1.** Real-time FFT and cutoff parameter control of low-pass filter



**Figure 2.** Real-time FFT and center frequency control of band-pass filter

We used non-blocking functions to let the user control the parameter of the filter through a graphical interface where we also showed in real-time an FFT of the buffer. We used the `scipy.fft(x)` function to compute the FFT. These first tests resulted in fluid stream reproduction. The module `scipy.signal` was then used to implement some digital filters of Scipy library by generating the coefficients of some Butterworth filters. We designed and tested low, high and band-pass filters with different orders (up to 7). For each filter, one or more parameters can be real-time controlled by the user. The interfaces created for low and band-pass filters are shown in Figures 1 and 2, respectively.

#### 3.2 Diode ring modulator

We implemented an algorithm that performs a digital simulation of the diode ring modulator [12] (Network Model showed in Figure 3) in order to stress the system. The first

trial developed using interpreted Python code suddenly revealed the main limit of this solution, i.e. execution time hundreds of times over the limits. After a brief analysis, we discovered that the main execution time was spent to compute the equations of the circuit which are complex operations performed several times for each sample. First of all, we tried some simple code optimizations (e.g. substitution of power functions with cascaded multiplications) but the amount of saved time was insufficient. Then, we tried to use Cython to compile the code; the only modification we made was to assign the types of the variables using the `cdef` construct. We left the structure and the variable scope belonging to Python unchanged. It was surprising how the performance increased: the average execution time reached the same order of the real-time execution. Eventually, in order to reach the real-time execution, we decided to translate the main 9 equations that describe the circuit in C language and bind the code with the cythonization process. As we can see in Table 1, this simple translation gives the necessary power to reach the real-time execution of the code. An example of the output of the system is shown in Figure 4.

```

=====
use_ring.py
=====
import c_ring
...
over_f = 30 #oversampling factor
T = 1./(s_rate*over_f)

b,a = butt_low(sample_rate/over_f, s_rate, 5)
zi = signal.lfilter_zi(b,a)

def process(c_in, m_in):
    global zi
    return c_ring.compute(c_in, m_in, CHUNK_SIZE,
                          T, over_f, a, b, zi)

=====
c_ring.pyx
=====
import cython
...
#declare interface to C code
cdef extern void ring(double* c,double* m,
                      int s, float T)

def compute(np.ndarray[double] c,
            np.ndarray[double] m,
            int s, float T, int over_f,
            np.ndarray[double] a,
            np.ndarray[double] b,
            np.ndarray[double] zi):

    # c-array that contains samples
    cdef double *cdata
    cdef double *mdata
    ...
    #resample buffer using 0-padding
    idx = np.repeat(np.arange(c.shape[0]+2),
                    (over_f-1) [1*(over_f-1):-1*(over_f-1)])
    c= np.insert(c, idx, 0)
    m=np.insert(m,idx,0)
    ...
    ring(cdata, mdata, s, T)
    ...

```

```

#low-pass filter applied
out, zi = signal.lfilter(b,a,c, zi=zi)
return out[::over_f], zi

=====
ring.c
=====

/*tecnical electrical data*/
const double C = 10e-9; // F
const double C_p = 10e-9 ; // F
const float L = 0.8; // H
const int R_a = 600; // ohm
const int R_i = 50; // ohm
const int R_m = 80; // ohm

double u1 = 0;
double u2 = 0;
double u3 = 0;
double u4 = 0;
double u5 = 0;
double u6 = 0;
double u7 = 0;

double i1 = 0;
double i2 = 0;

void ring(double* carr_in, double* mod_in,
          int s, float T){
    int i=0;
    for (i = 0; i < samples; i++) {

        u1 = u1 + T/C*(i1 - (u4*u4*u4*u4*0.17)/2 +
            (u7*u7*u7*u7*0.17)/2 + (u5*u5*u5*u5*0.17)/2 -
            (u6*u6*u6*u6*0.17)/2 - (u1-mod_in[i]*0.2)/R_m );

        u2 = u2 + T/C*(i2 + (u4*u4*u4*u4*0.17)/2 +
            (u7*u7*u7*u7*0.17)/2 - (u5*u5*u5*u5*0.17)/2 -
            (u6*u6*u6*u6*0.17)/2 - u2/R_a );

        u3 = u3 + T/C_p*((u4*u4*u4*u4*0.17) -
            (u7*u7*u7*u7*0.17) + (u5*u5*u5*u5*0.17) -
            (u6*u6*u6*u6*0.17)- (u3)/R_i );

        u4 = u1/2 - u3 - carr_in[i]*0.2 - u2/2;
        u5 = -u1/2 - u3 - carr_in[i]*0.2 + u2/2;
        u6 = u1/2 + u3 + carr_in[i]*0.2 + u2/2;
        u7 = -u1/2 + u3 + carr_in[i]*0.2 - u2/2;

        i1 = i1+ T/L*(-u1);
        i2 = i2+ T/L*(-u2);
        carr_in[i] = u2;
    }
    return;
}

```

### 3.3 VCS3

We further developed the implementation of the VCF digital simulation [13] to test the structure and the amount of common functions that could be used in different algorithms. In this case we confirmed that the time and the effort needed to translate an existing Matlab implementation (designed to work on a fixed array of samples) in real-time

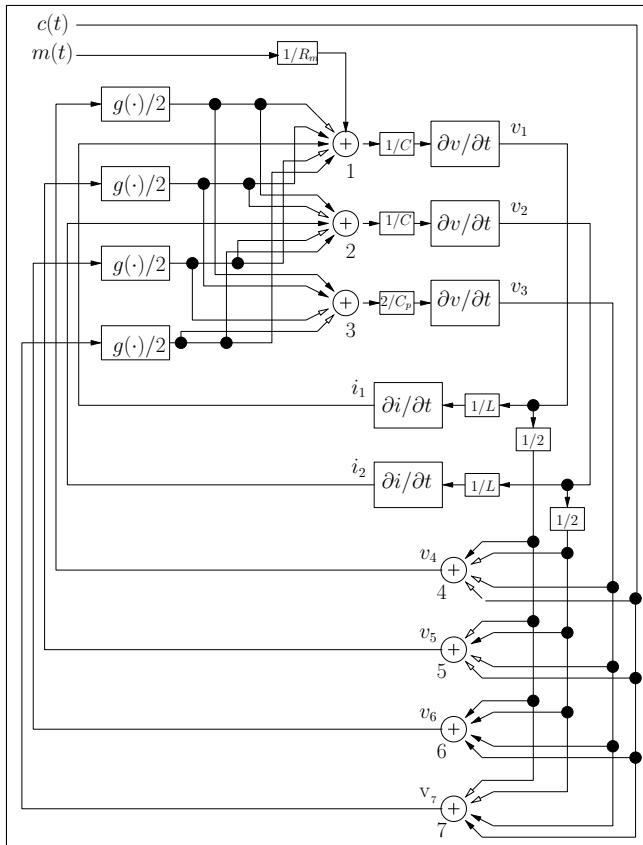


Figure 3. Network Model of the Diode Ring Modulator

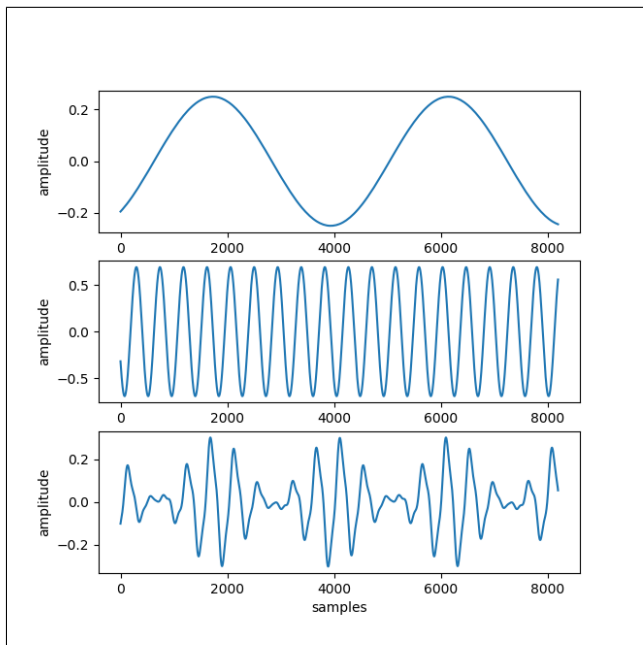


Figure 4. Time domain analysis of diode ring modulator (input-carrier: 10Hz, input-modulator: 100Hz, output) over 8192 samples

|                 | Time         |
|-----------------|--------------|
| Pyhton          | 1830 $\mu$ s |
| Cython          | 54 $\mu$ s   |
| Real-time limit | 22.6 $\mu$ s |
| Cython + C      | 9.2 $\mu$ s  |

Table 1. Mean execution time of a sample computed by Diode Ring Modulator over 1000 chunks at 44100 Hz.

|                 | Time         |
|-----------------|--------------|
| Pyhton          | 78.3 $\mu$ s |
| Real-time limit | 22.6 $\mu$ s |
| Cython          | 20.3 $\mu$ s |
| Cython + C      | 0.64 $\mu$ s |

Table 2. Mean execution time of a sample computed by VCS3 algorithm over 1000 chunks at 44100 Hz.

code is limited. Concerning all the code parts related to Python, only little modifications were done from the previous project while the C translation of the core algorithm was trivial. In the following code an example of the program structure is given: we can see how the compiled code is treated as a library while the Cython file contains only a wrapper of the C function and transformations of data types. In Table 2 we can see that both the Cython and C implementation are suitable for the real-time execution while the Python version remains over the limit. The Network Model and the frequency domain analysis with white noise input are shown in Figure 5 and 6, respectively.

```

=====
use_vcf.py
=====
import c_vcf
...
T = 1./sample_rate
def process(data):
    return vcf.compute(data,CHUNK_SIZE, T)

=====
c_vcf.pyx
=====
import cython
...
#declare interface to C code
cdef extern void vcf(double* d,
                    int s, float T)

def compute(np.ndarray[double] d,
            int s, float T):
    # c-array that contains samples
    cdef double *cdata
    ...
    vcf(cdata, s, T)
    ...
    return d
    
```

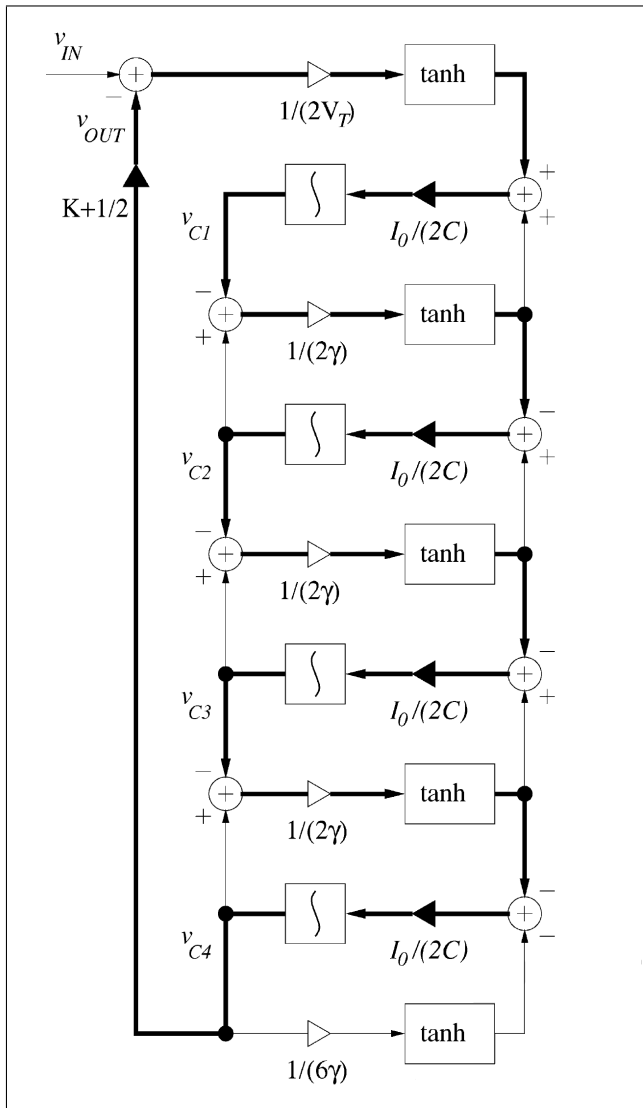


Figure 5. Network Model of the VCF

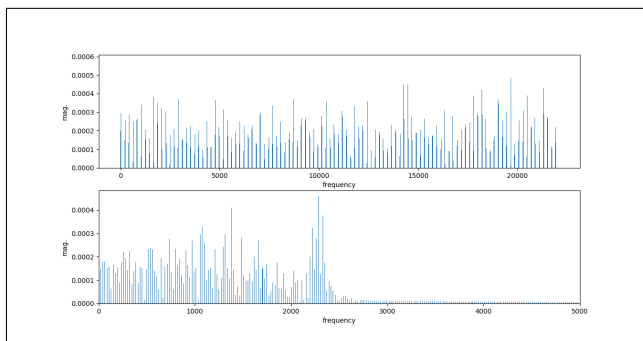


Figure 6. Frequency domain analysis of VCS3 with white noise input

```

=====
vcf.c
=====
#include <stdio.h>
#include <stdlib.h>
#include <math.h>

double gfdbk=5.175;
float I0= 30e-6;
const float err= 10e-4;
...
void vcf(double* vin, int samples, float T){
    int i=0;
    for(i=0;i<samples;i++){
        float vc4T=0;
        float vc4Past=1;
        int nIter=0;

        while(abs(vc4T-vc4Past)>err*abs(vc4Past)){
            vc4Past=vc4T;
            vin1T= tanh((vin[i] - vout)/2*Vt);

            //block 1
            xc1T = (I0/2/c)* (vin1T+vc11T);
            vc1T = T/2*xc1T + s1;
            vc11T = tanh((vc2T-vc1T)/(2*gmma));

            //block 2
            xc2T = (I0/2/c)* (vc21T+vc11T);
            vc2T = T/2*xc2T + s2;
            vc21T = tanh((vc3T-vc2T)/(2*gmma));

            //block 3
            xc3T = (I0/2/c)* (vc31T+vc21T);
            vc3T = T/2*xc3T + s3;
            vc31T = tanh((vc4T-vc3T)/(2*gmma));

            //block 4
            xc4T = (I0/2/c)* (-vc41T-vc31T);
            vc4T = T/2*xc4T + s4;
            vc41T = tanh((vc4T/(6*gmma));

            vout = vc4T/2 + vc4T*gfdbk;

            nIter = nIter+1;
            if (nIter == 100){
                nIter=0;
                break;
            }
        }
        ...
    }
}

```

Through the development of these Virtual Analog models the potentialities of the framework have been stressed out. As we can see from the time measurements, the chances to use Python for audio real-time application are limited but the usage of Cython and the C allows to reduce the execution time of the critical part of the code. As a final result the developed code belongs to separated modules with different execution speed:

- Python interpreted code can be used for environment setup, data conversions, signal analysis and plots (slow speed)
- Cython code can be used for re-sample, memory allocation, simple filters (type declaration for variables is provided)
- C code with Python bindings can be used for low-level code like arithmetical operations (high speed)

#### 4. CONCLUSION

We approached Python as a tool for the development of real-time sound synthesis procedures. With this language we developed a software structure made of three layers embracing different programming levels, depending on the efficiency needed by the application. Several digital filters and effects can be connected together in a simple way. In fact, we encapsulated each filter in a Python module that can be either used as a black box or manipulated by the user.

We also found an effective methodology to measure performances, and hence discover critical sections of the code which harm real-time execution. These sections can be re-encapsulated in the lower levels, with consequently dramatic increase of the computational speed. Future work will introduce the Qt technology in the existing framework, to create more rich user interfaces. We are also planning to realize a larger library of Virtual Analog models and running examples, whose aim will be to help novice users program sound processing procedures at different levels of granularity however through a widely used programming language, renowned for its flexibility and immediacy of use.

#### 5. REFERENCES

- [1] P. Guo, "Python is now the most popular introductory teaching language at top u.s. universities.." <https://cacm.acm.org/blogs/blog-cacm/176450-python-is-now-the-most-popular-introductory-teaching-language-at-top-u-s-universities/fulltext>, 2014.
- [2] C. G. Mason1, M. and M. Raadt, "Trends in introductory programming courses in australian universities c languages, environments and pedagogy," in *Australasian Computing Education Conference (ACE2012)* (M. Raadt and A. Carbone, eds.), vol. 123 of *CRPIT*, (Melbourne, Australia), pp. 33–42, ACS, 2012.
- [3] A. B. Downey, *Think DSP: Digital Signal Processing in Python*. O'Reilly Media, Inc., 1st ed., 2016.
- [4] S. D. Krishan Kumar, "Programming languages: A survey," *International Journal on Recent and Innovation Trends in Computing and Communication (IJRITCC)*, pp. 307 – 313, 05 2017.
- [5] S. Behnel, R. Bradshaw, C. Citro, L. Dalcin, D. S. Seljebotn, and K. Smith, "Cython: The best of both worlds," *Computing in Science and Engineering*, vol. 13, pp. 31–39, March 2011.
- [6] Sebastián Sepúlveda, Pablo Reyes, and Alejandro Weinstein, "Visualizing physiological signals in real-time," in *Proceedings of the 14th Python in Science Conference* (Kathryn Huff and James Bergstra, eds.), pp. 190 – 194, 2015.
- [7] Brian McFee, Colin Raffel, Dawen Liang, Daniel P.W. Ellis, Matt McVicar, Eric Battenberg, and Oriol Nieto, "librosa: Audio and Music Signal Analysis in Python," in *Proceedings of the 14th Python in Science Conference* (Kathryn Huff and James Bergstra, eds.), pp. 18 – 25, 2015.
- [8] D. Bellini, "Expressive digital signal processing (dsp) package for python." <https://github.com/danilobellini/audiolazy>, 2016.
- [9] O. Blanger, "Python dsp module." <https://github.com/belangeo/pyo>, 2018.
- [10] H. Pham, "Pyaudio website." <https://people.csail.mit.edu/hubert/pyaudio/>, 2006.
- [11] R. Kern, "Line profiler code repository." [https://github.com/rkern/line\\_profiler](https://github.com/rkern/line_profiler), 2017.
- [12] F. Fontana and E. Bozzo, "Explicit fixed-point computation of nonlinear delay-free loop filter networks," *IEEE/ACM Transactions on Audio, Speech and Language Processing*, vol. 26, pp. 1884–1896, Oct 2018.
- [13] F. Fontana and M. Civolani, "Modeling of the ems vcs3 voltage-controlled filter as a nonlinear filter network," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, pp. 760–772, May 2010.

# IL GIOIELLO ELETTRACUSTICO

**Arcangelo Di Donato**

Conservatorio “C. Pollini”, Padova  
arcadido@gmail.com

**Alberto Novello**

Conservatorio “C. Pollini”, Padova  
albynovello@gmail.com

**Giorgio Klauer**

Conservatorio “B. Marcello”, Venezia  
giorgio@klauer.it

## ABSTRACT

Il gioiello elettroacustico è un piccolo circuito elettronico indossabile che emette o è azionato dal suono. In questo articolo viene introdotto un ambito di ricerca e sviluppo di gioielli che allo sfoggio dell'estetica aggiungono qualità interattive, estendendone la fruizione al senso dell'udito.

## 1. INTRODUZIONE

Nella pubblicistica, quando si parla di *wearable design* (progettazione di oggetti indossabili) ci si riferisce solitamente alla dimensione tecnologica dei *wearable electronics* [1]. Tuttavia, gli oggetti tecnologici indossabili hanno origini antiche: già nel 1500 l'inventore tedesco Peter Henlein produsse il primo orologio da portare al collo come una collana. È con la diffusione della microelettronica su larga scala, negli anni '80 del Novecento, che questa tipologia di oggetti si diffonde universalmente - e l'orologio meccanico si evolve in orologio-calcolatrice come il Casio Databank; l'attuale assetto dei *wearable* si definisce negli anni Duemila grazie alla ulteriore miniaturizzazione e l'implementazione dei sensori e della comunicazione wireless - si confronti l'uscita nell'aprile 2015 dell'Apple Watch. Gli oggetti indossabili elettronici sono oggi molto diffusi, anche se sono perlopiù legati a una singola funzionalità come il monitoraggio delle attività biologiche per il fitness (*activity trackers*), la geolocalizzazione e la comunicazione (nell'ambito dell'*internet of things*). Per queste protesi elettroniche<sup>1</sup> viene richiesto un design funzionale e una immediata intelligibilità.<sup>2</sup>

### 1.1 Il gioiello elettronico

Il gioiello con componenti elettroniche fa parte dei *wearable electronics* ma ha caratteristiche peculiari rendendo obliquo il discorso su funzione ed ergonomia. Conviene pertanto definire l'ambito progettuale del gioiello elettronico. Questo deve avere in primis le caratteristiche del gioiello, cioè essere svincolato da necessità e bisogni pratici.

<sup>1</sup> “I nostri corpi sono la nostra principale interfaccia con il mondo. I sistemi interattivi [...] si trovano vicino alla nostra pelle, nei nostri vestiti, e a volte li sentiamo come parte di noi” [2].

<sup>2</sup> Cfr. ad es. <https://designguidelines.withgoogle.com/wearos/wear-os-by-google/creative-vision.html>

Copyright: ©2018 Arcangelo Di Donato et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution License 3.0 Unported](https://creativecommons.org/licenses/by/3.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

“Gioiello” deriva dal francese antico *joel*<sup>3</sup> che è il latino *ioalis*, aggettivo di *ioocus*: scherzo, gioco. Possiamo quindi proporre che al gioiello appartengano le qualità che Roger Caillois ha definito per il gioco [3].

In questo parallelismo ritroviamo che il gioco è riconducibile a un'attività “libera” (un giocatore non può essere obbligato, come libera deve essere la scelta di indossare o meno un ornamento), “separata” (circoscritta entro precisi limiti di tempo e di spazio fissati in anticipo, che nel caso del gioiello vuol dire l'essere indossato in occasioni definite), “incerta” (lo svolgimento e il risultato non possono essere conosciuti in anticipo, è il caso della discrezionalità nel giudizio estetico del gioiello sfoggiato), “improduttiva” (non crea beni o ricchezza), “regolata” (risponde a regole proprie che inibiscono le leggi codificate) e “fittizia” (il giocatore è consapevole delle differenze con la vita reale).<sup>4</sup>

Definito l'ambito del gioiello, definiamo lo spazio circoscritto al gioiello elettronico: il gioiello elettronico è un gioiello con componenti elettroniche attive che ne qualificano l'estetica.

### 1.2 Gioielli elettronici in vendita

I gioielli elettronici in vendita usano in particolare i led per aumentare la visibilità e la variazione cromatica dell'oggetto. Un esempio diffuso nelle bigiotterie sono gli orecchini luminosi. Tradizionalmente il compito di punto luce viene assolto dai diamanti, ma nel caso di orecchini luminosi questa funzione viene svolta da led che possono produrre luce bianca o, con led RGB, avere un colore mutevole. Ne sono un esempio la produzione della Lumen Electronic Jewelry.<sup>5</sup> Il negozio online propone gioielli con led azionati da un microcontrollore talvolta alimentati da energia solare. Uno degli esempi più recenti di oggetti indossabili che rientra nell'ambito dei *fashionable wearable*<sup>6</sup> sono gli NFC ring.<sup>7</sup> Questi anelli rispondono ai bisogni di un'estetica che non sovrasti la funzione, e la funzione è esaltata dall'evidenza estetica del circuito NFC. Ringly<sup>8</sup> produce anelli che avvisano in maniera discreta, con luci o vibrazioni, la ricezione di messaggi da diverse piattaforme social. Una vera novità nella gioielleria è l'in-

<sup>3</sup> <http://www.treccani.it/vocabolario/gioiello>

<sup>4</sup> “Il gioco è occasione di puro dispendio: di tempo, di energie, d'intelligenza, di abilità e a volte di denaro, per l'acquisto degli accessori del gioco o per pagare eventualmente l'affitto del locale.” [3]

<sup>5</sup> <http://www.lumenelectronicjewelry.com>

<sup>6</sup> *Fashionable wearable* sono “designed garments and accessories that combines aesthetics and style with functional technology” [4].

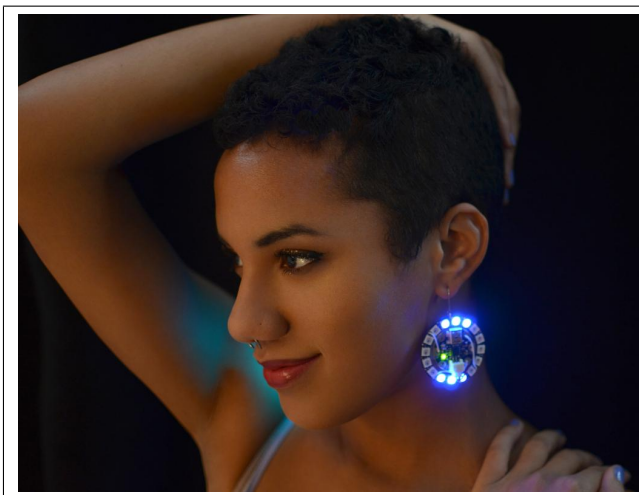
<sup>7</sup> <http://store.nfcring.com>

<sup>8</sup> <https://ringly.com>

produzione dei display flessibili e-ink. La L!ber8<sup>9</sup> propone il bracciale Tago Arc il cui display può cambiare il disegno decorativo comunicando via NFC con uno smartphone.

### 1.3 Gioielli elettronici di maker

La diffusione di piattaforme di prototipazione e microcontrollori facilmente programmabili, i costi ridotti di sensori e attuatori hanno portato la comunità dei maker a espandersi fino a toccare il campo del gioiello. Nel gioiello si è riversato lo spirito libero e creativo di chi cerca l'alternativa a sistemi chiusi appannaggio di poche grandi aziende. In rete si trovano molti esempi di prototipi autoconstruiti come i Reactive Jeweleries di Maxim Krentovskiy.<sup>10</sup> Questi gioielli hanno una estetica vistosa e nascondono al loro interno un microcontrollore ATtiny85 che aziona dei led in base a sensori di luce o calore. Altri ricercatori cercano di individuare il punto di incontro tra le qualità tattili, percettive e psicologiche del gioiello tradizionale con l'“aumento” di alcune qualità sensoriali grazie alla tecnologia: è il caso del progetto Wearable Sensor Network [5]. Un altro esempio, dal quale si è partiti per una realizzazione discussa più avanti, sono i GEMMA Hoop Earrings (Figura 1),<sup>11</sup> orecchini con luminosità e colori variabili basati sulla scheda Gemma M0 e un NeoPixel Ring.



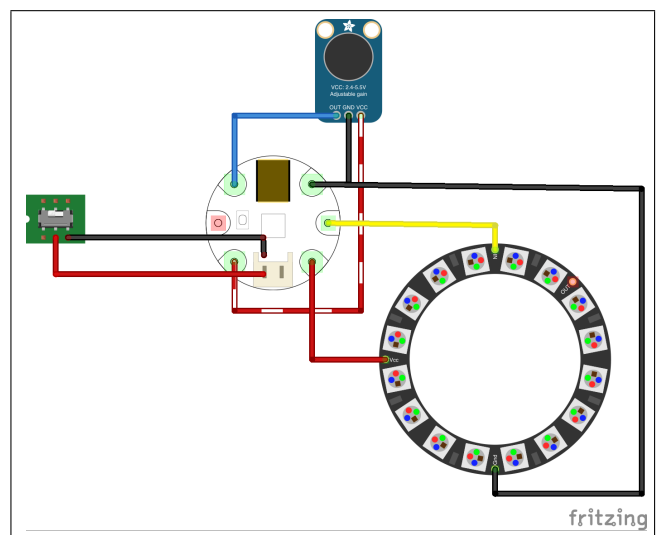
**Figura 1.** GEMMA Hoop Earrings. © <http://www.adafruit.com>

### 1.4 Il gioiello elettroacustico

Se il gioiello elettronico è poco diffuso, la sua declinazione nel gioiello elettroacustico, ossia con un'elettronica e un design che ne esalta le qualità non solo visive e tattili ma anche acustiche, è un terreno pressoché vergine. Al momento non risulta alcuna produzione di gioielli elettroacustici. La comunità dei maker propone però prototipi come il Neopixel VU earrings<sup>12</sup> che riprende il proget-

to di GEMMA Hoop Earrings con l'aggiunta di un microfono e un VU meter nella forma di un Neopixel Ring (Figura 2). Un importante riferimento per lo sviluppo di progetti con microcontrollori come l'ATtiny85 è Technoblogy.<sup>13</sup> Technoblogy non si occupa di gioielli, ma i progetti proposti per l'audio possono tornare di utile riferimento nello sviluppo di gioielli elettroacustici. I progetti di interesse ad oggi proposti sono:

- Audio Sample Player
- Simple Tones for ATtiny
- Waveform Generation using an ATtiny85
- Tiny Synth
- ATtiny85 Sound Level Meter
- Digital Music Box
- A Lightweight Alternative to tone
- Audio Pitch Shifter
- Tiny Function Generator
- Tiny Function Generator Sine Wave
- Harmonic Function Generator



**Figura 2.** Schema elettrico degli orecchini con VU meter.

## 2. SUONARE CON L'ATTINY85

La scelta del microcontrollore è stata cruciale per definire tutti i componenti compatibili. Ci si è indirizzati verso il microcontrollore Atmel ATtiny85 per le dimensioni contenute, un adeguato numero di pin, velocità di calcolo e disponibilità di alcune librerie dedicate che rendono compatibili sensori utili ai progetti. Per il risparmio energetico ci si è orientati verso la versione del microcontrollore a basso voltaggio, l'ATtiny85V, che funziona anche a 1,8 V fino a 4MHz, ma che può lavorare, aumentando voltaggio e consumi, fino a 10 MHz (contro i 20MHz dell'ATtiny85). Anche l'ATtiny85, come gran parte delle schede Arduino compatibili, permette di emettere segnali sonori. Oltre

<sup>9</sup> <https://www.liber8tech.com>

<sup>10</sup> <https://hackaday.io/project/28396-reactive-jeweleries>

<sup>11</sup> <https://learn.adafruit.com/gemma-hoop-earrings>

<sup>12</sup> <https://blog.adafruit.com/2014/08/13/neopixel-vu-earrings-wearablewednesday>

<sup>13</sup> <http://www.technoblogy.com>

ad onde quadre il microcontrollore può generare onde più complesse aggiungendo armoniche.<sup>14</sup> Inoltre può riprodurre campioni a 8 bit<sup>15</sup> e suonare polifonie.<sup>16 17</sup> Tuttavia, le alternative alla monofonia con onde quadre richiedono un'amplificazione che si vuole evitare per motivi di contenimento dei consumi e degli ingombri dell'hardware. Le scelte compositive saranno quindi orientate alla monofonia con onde quadre. Le funzioni per generare suoni sono diverse e la convenienza dell'una o dell'altra dipende anche dal core utilizzato per la programmazione. L'ATtiny85 può essere programmato con vari core che usano pin e counter diversi per svolgere alcune funzioni. Ad esempio il core ATtiny di David Mellis<sup>18</sup> fornisce assegnazioni di pin in modo che le routine di Arduino possano essere utilizzate con ATtiny85 ma non supporta la funzione `tone()`. Il core Arduino-Tiny<sup>19</sup> fornisce routine di core per `tone()` ma solo a 8MHz. Si sono prese in considerazione diverse funzioni per generare suoni:

- `tone()`<sup>20</sup>
- `beep()`<sup>21</sup>
- `TrinketTone()`<sup>22</sup>
- `playTone()` [6]
- `TinyTone()`<sup>23</sup>
- `note()`<sup>24</sup>

Le ultime due sono eseguite anche a 1MHz e sono le candidate ideali per progetti che richiedono velocità di calcolo e bassi consumi.

### 3. PROGETTI

Vengono qui proposte quattro idee di gioielli elettroacustici in diverso stato di sviluppo.

#### 3.1 Pendente

Il primo gioiello progettato è un prototipo di pendente (Figura 3). I componenti elettronici sono:

- Arduino Nano
- Altoparlanti miniaturizzati piezoceramici cod. 491768<sup>25</sup>
- Catenine metalliche che fungono da resistenze variabili
- Batteria 3,7 V ( 150 mAh)

<sup>14</sup> <http://www.technoblogy.com/show?QVN>

<sup>15</sup> <http://www.technoblogy.com/show?QBB>

<sup>16</sup> <http://www.technoblogy.com/show?Q7H>

<sup>17</sup> <http://www.technoblogy.com/show?11TQ>

<sup>18</sup> <https://github.com/damellis/attiny>

<sup>19</sup> <https://code.google.com/archive/p/arduino-tiny>

<sup>20</sup> <https://www.arduino.cc/reference/en/language/functions/advanced-io/tone>

<sup>21</sup> [https://msdn.microsoft.com/it-it/library/windows/desktop/ms679277\(v=vs.85\).aspx](https://msdn.microsoft.com/it-it/library/windows/desktop/ms679277(v=vs.85).aspx)

<sup>22</sup> <http://w8bh.net/avr/TrinketTone.pdf>

<sup>23</sup> <http://www.technoblogy.com/show?KVO>

<sup>24</sup> <http://www.technoblogy.com/show?20MO>

<sup>25</sup> <http://www.conrad.it/ce/it/product/491768/Speaker-in-miniatura-Pressione-acustica-92-dB--491768-1-pz>



Figura 3. Il gioiello pendente.

Il pendente presenta tre catenine, quella centrale è collegata all'alimentazione, le altre agli ingressi analogici del microcontrollore. La variazione di resistenza dovuta al contatto tra i pendenti offre dati che possono essere elaborati per definire, ad esempio, il pitch e la sequenza di suoni. I materiali del gioiello sono stati stampati da <http://www.shapeways.com>: la scatola è in *sandstone*, i "tentacoli" in plastica. È disponibile un video che illustra il gioiello in funzione: <https://youtu.be/tpKXFVZ-bI>.

#### 3.1.1 Evoluzione del progetto

Il pendente presenta tre altoparlanti che, essendo vicini, non consentono la spazializzazione del suono. Altro problema sono le dimensioni del microcontrollore: benché l'Arduino Nano sia molto versatile e dalle dimensioni contenute, è in ogni caso sovradimensionato come ingombro e consumi. L'alternativa è utilizzare solo il chip del microcontrollore, ad esempio l'ATtiny85 (Figura 4). Il progetto si è quindi evoluto con i seguenti componenti:

- ATtiny85V-10SU
- Pannello solare 40mA silicio monocristallino N
- BQ25504 Solar Cell LiPo Charger
- Sparkfun Inductive Charging Set
- Batteria 3,7 V (100 mAh)
- Altoparlanti miniaturizzati piezoceramici cod. 491768

- RGB led Neopixel
- Lithium Battery Charger Module
- Tilt Switches Rolling Ball
- Resistenze
- Catenine

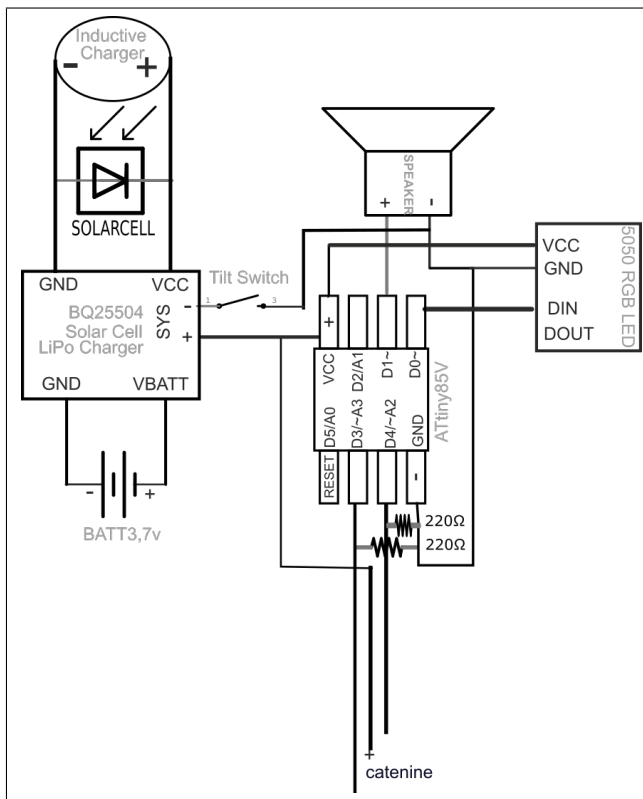


Figura 4. Schema elettrico del pendente.

Il pendente sfrutta l'energia solare per caricare la piccola batteria. Sotto il pannello solare è celato un sistema di ricarica a induzione. Quando il gioiello è posto sulla basetta ad induzione di ricarica, il tilt switch spegne l'oggetto. Bozza del progetto e codice sono disponibili al sito <https://demoarduino.blogspot.com/2017/11/gioiello-e-contatti.html>.

### 3.2 Anello

La progettazione di un anello elettroacustico (Figura 5) ha comportato la selezione di componenti di dimensione adeguata per essere nascosti dallo speaker messo in evidenza e a caratterizzazione dell'estetica del gioiello (Figura 6). I componenti utilizzati sono:

- ATtiny85-20SU
- Altoparlante miniaturizzato piezoceramico cod. 491768
- Accelerometro ADXL345
- Lithium Battery Charger Module
- Batteria 3,7 V (100 mAh)

L'ATtiny85 sfrutta i valori rilevati dall'accelerometro per determinare pitch, sequenza e logica dell'articolazione dei suoni. Un prototipo è attualmente in fase di elaborazione.

### 3.3 Orecchini

Il progetto riprende l'esempio Neopixel VU earrings, con delle modifiche. All'Arduino Gemma si preferisce il più piccolo ATtiny85. Inoltre, si fa a meno dell'amplificatore MEMS Microphone Breakout sfruttando alcune proprietà dell'ATtiny85 per la conversione da analogico a digitale.

Il segnale captato dal microfono viene digitalizzato utilizzando uno dei convertitori dell'ATtiny85. La modalità di conversione analogica single-ended standard offre una sensibilità massima di 1,1 V utilizzando il riferimento di tensione interno da 1,1 V per i 12 bit, non compatibile con il segnale del microfono. L'approccio abituale è di aggiungere un amplificatore operazionale (come il MEMS Microphone Breakout nel progetto citato).<sup>26</sup>

L'ATtiny85 offre un'opzione di guadagno x20, corrispondente a una sensibilità su scala reale di 55 mV, ideale per questa applicazione. Il guadagno x20 è disponibile solo nelle modalità di conversione analogica differenziale per la quale vengono impiegati due ingressi analogici. Nel caso non ci fosse disponibilità di un altro ingresso analogico, si può configurare il pin di Reset come ADC0 aggiungendo un offset (il Reset si attiva sotto i 2,5 V). L'ADC è dunque configurato in setup per la conversione differenziale usando ADC0 (PB5, il Reset) e ADC1 (PB2) con un guadagno x20. L'ADC è pilotato da un clock separato dal clock di sistema. È necessario utilizzare resistenze con margine di errore 1% o minore.

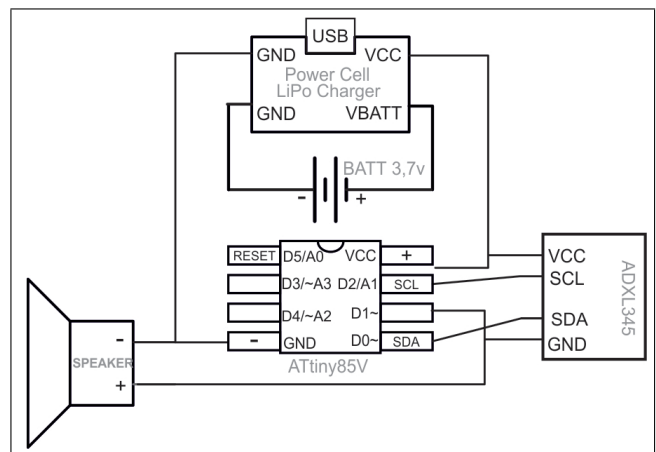


Figura 5. Schema elettrico dell'anello.

### 3.4 Bracciale "Voodoo Synth"

Esistono numerosi progetti che usano il microcontrollore ATtiny85 per costruire sintetizzatori. Quello più interessante è forse "8bit Mix Tape",<sup>27</sup> di cui esistono diverse varianti. Partendo dalle esperienze già svolte si è cercata una via alternativa perché tutti i progetti già sviluppati richiedono una amplificazione esterna, cosa che si voleva evitare. Il "Voodoo Synth" presenta una serie di ingressi selezionabili infilando uno spillo collegato al microcontrollore con un filo da sartoria conduttivo. Attraverso

<sup>26</sup> La soluzione per evitare l'amplificatore esterno è stata presa dal sito <https://hackaday.io/project/28396-reactive-jeweleries>.

<sup>27</sup> [http://wiki.sgmk-ssam.ch/wiki/8bit\\_Mix\\_Tape](http://wiki.sgmk-ssam.ch/wiki/8bit_Mix_Tape)



**Figura 6.** Mockup dell'anello elettroacustico: lo speaker diventa elemento qualificante dell'estetica.

gli ingressi si possono configurare: preset per i sequencer, beat e timbro dei suoni. Il bracciale al momento è poco più di un concept (Figura 7). Una bozza del progetto e del codice è disponibile al sito <https://demoarduino.blogspot.com/2018/04/woodoosnth.html>.

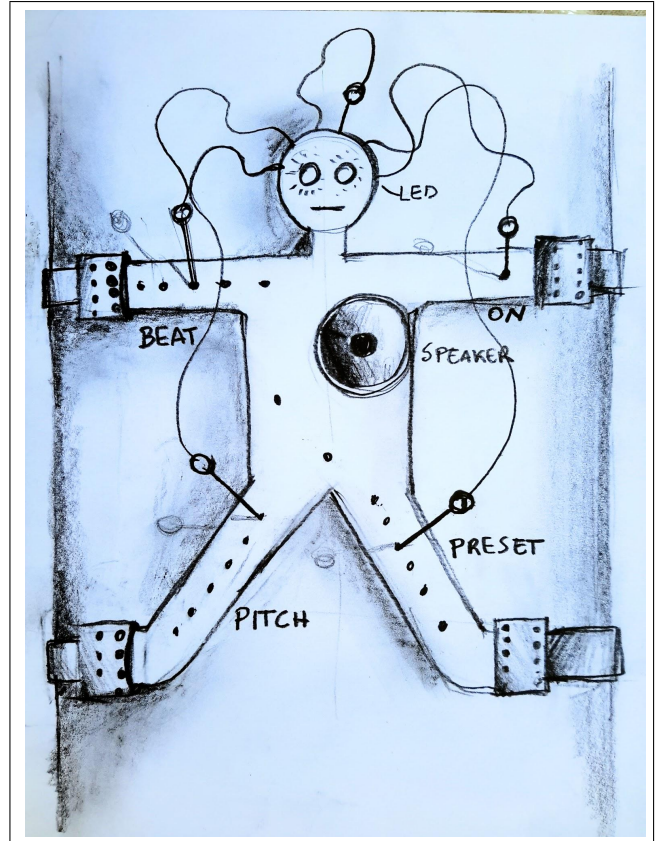
#### 4. CONCLUSIONI

Nel presente articolo si sono affrontati i problemi tecnici che riguardano la creazione del gioiello elettroacustico individuando delle possibili direzioni per lo sviluppo di nuove idee. In particolare ci si è soffermati sui problemi dell'hardware. Tuttavia, le questioni da risolvere sono tutte interconnesse: tensione e amperaggio dell'alimentazione sono legati al settaggio del clock interno e della CPU del microcontrollore, alle richieste energetiche di sensori e attuatori e non ultime agli ingombri delle batterie.

Poiché la dimensione artigianale è caratteristica dell'oreficeria qualificata, pensare ad un approccio di autoproduzione appare scontato. Tuttavia, al momento si vogliono esplorare soltanto le possibilità di aprire nuovi scenari. Tra le difficoltà riscontrate c'è la miniaturizzazione, che per un'azienda elettronica è su un altro livello rispetto a quello di un maker. Questo obbliga a ricercare soluzioni creative accettando compromessi che caratterizzeranno inevitabilmente l'estetica dell'oggetto. Proseguendo la ricerca dovrà essere sviluppata maggiormente l'estetica artistica e musicale, sulla quale non ci si è ancora soffermati abbastanza per qualificare la valenza creativa nell'ambito del gioiello elettroacustico.

#### 5. BIBLIOGRAFIA

[1] K. Tehrani and A. Michael, "Wearable technology and wearable devices: Everything you need to know," *Wearable Devices Magazine*, Marzo 2014. <http://www.wearabledevices.com/what-is-a-wearable-device>.



**Figura 7.** Bozzetto del bracciale "Voodoo Synth".

[2] K. Hartman, *Elettronica wearable. Disegna, prototipa e indossa i tuoi vestiti e accessori interattivi*. Milano: LSWR, 2015.

[3] R. Caillois, *I giochi e gli uomini: La maschera e la vertigine*. Milano: Bompiani, 1981. Tit. orig. *Les Jeux et les hommes: le masque et le vertige*, Gallimard, 1958.

[4] S. Seymour, *ashionable Technology: The intersection of design, fashion, science and technology*. Springer, 2008.

[5] A. L. Ju and M. Spasojevic, "Smart jewelry: The future of mobile user interfaces," in *Proceedings of the 2015 Workshop on Future Mobile User Interfaces (FutureMobileUI@MobiSys)*, (Firenze), pp. 13–15, 2015.

[6] M. Barela, *Getting Started with Adafruit Trinket: 15 Projects with the Low-Cost AVR ATtiny85 Board*. Maker Media, 2014.

## Distorsione di Massa

**Stefano Catena**

Conservatorio Giuseppe Verdi Milano  
stefano.catena23@gmail.com

**Andrea Bolzoni**

Conservatorio Giuseppe Verdi Milano  
andreabolzonimusic@gmail.com

### ABSTRACT

Distorsione di Massa é un'installazione interattiva che, sotto forma di gioco, permette all'utente di riflettere sul tema delle fake news. Si presenta come una postazione di ascolto, costituita da una sedia e da un tavolo su cui si trova un tablet, posizionata al centro di uno spazio di diffusione sonora quadrifonica. Il gioco é organizzato in livelli, di difficoltà crescente, in cui l'obiettivo é quello di riconoscere una frase musicale che rappresenta una notizia vera. All'inizio di ogni livello viene fatta ascoltare la frase musicale "vera" che, successivamente, viene moltiplicata e le cui copie vengono sottoposte a un processo di distorsione. Passato un lasso di tempo direttamente proporzionale alla difficoltà del livello, viene sbloccata la possibilità di selezionare, attraverso la visualizzazione delle posizioni delle sorgenti sul tablet, quale tra le voci ascoltate é quella rimasta inalterata.

### 1. INTRODUZIONE

Il dibattito riguardante la qualità delle informazioni che al giorno d'oggi vengono cercate o subite attraverso l'utilizzo di internet, e più in specifico dei social network, é di grande rilevanza.

Un importante studio effettuato dall'MIT e pubblicato nel 2018 su Science [9] analizza in maniera organica il fenomeno delle *fake news* descrivendo in maniera approfondita le caratteristiche della loro diffusione attraverso i *social network*. Gli aspetti più rilevanti riguardano la maggior velocità di diffusione delle notizie false rispetto alle notizie vere e il motivo per cui queste si diffondono più rapidamente: gli utenti, in generale, sono maggiormente attratti dalle novità e quindi tendono a condividerle e a diffonderle con facilità.

Al contrario, lo studio del 2017 *The Spread of Fake News by Social Bots* [7] sostiene che, analizzando diversi milioni di messaggi diffusi su Twitter durante la campagna elettorale per le presidenziali del 2016 negli Stati Uniti, si evince il ruolo chiave dei *bots* nella diffusione di disinformazione. In accordo con un altro studio, sempre del 2017, *Social Media and Fake News in the 2016 Election* [1], vengono evidenziate le due principali ragioni che incentivano l'esistenza dei siti di fake news: la prima consiste nei ricavi ottenuti dalle pubblicità derivanti dall'alto numero di

click e la seconda é la possibilità di influenzare l'opinione politica dell'elettorato.

Il 62% degli adulti che vivono negli Stati Uniti si tengono informati attraverso i social media, tra cui Facebook, piattaforma su cui, nel periodo delle elezioni americane del 2016, sono state diffuse più *fake news* che storie vere. *Fake news* che molte persone dicono di aver preso per vere nel momento in cui le hanno lette [1].

Ciò che ne risulta sono l'atteggiamento di superficialità con cui queste notizie vengono spesso affrontate e diffuse (condivise, diremmo, parlando di *social network*) e il potenziale livello di rischio globale dell'imponente diffusione delle *fake news* [7]. Queste circostanze hanno alimentato la spinta ad approfondire l'argomento con l'obiettivo di creare uno strumento, una macchina sonora, in grado di sensibilizzare e di far riflettere.

Il progetto Distorsione di Massa nasce dall'idea di rappresentare sotto forma sonora il processo di distorsione del significato di un fatto considerato oggettivamente vero nel suo contenuto (*fake news* é "distorsione", non "filtraggio" [1]).

La prima fase di analisi é consistita nell'individuare:

- la sorgente da cui attingere un testo rappresentante un fatto considerato vero;
- la tecnica di codifica del testo in suono;
- la rappresentazione metaforica del processo di distorsione.

La difficoltà nel sistematizzare una ricerca di notizie che contenesse anche un controllo della loro veridicità si é risolta nella scelta di individuare, in fase di scrittura del codice, alcuni titoli di notizie facilmente verificabili da implementare nel sistema.

Per la codifica del testo in suono si é deciso di associare ogni carattere a una combinazione *pitch-ritmo*.

Il processo di distorsione viene rappresentato su due livelli:

- variazione della frase cambiando le singole parole con dei sinonimi;
- variazione del contenuto spettrale del suono, l'articolazione dell'esecuzione o del pitch.

### 2. BACKGROUND

L'utilizzo di lettere dell'alfabeto per la notazione musicale é rintracciabile fin dal VI secolo, quando il filosofo Boezio associó le prime quattordici lettere dell'alfabeto latino alle altezze musicali. Successivamente, all'inizio dell'XI secolo, Guido d'Arezzo, per aiutare l'apprendimento de-

gli intervalli dell'esacordo musicale, ideò un sistema che associava le prime sillabe dei versi dell'Inno a San Giovanni alle altezze corrispondenti, ponendo le basi del sistema da cui ancora oggi derivano i nomi delle note che comunemente vengono utilizzati in una parte degli stati europei.

Basandosi sul sistema della solmisazione<sup>1</sup>, il compositore Josquin Desprez adottò per primo una tecnica, che verrà successivamente definita da Zarlino *soggetto cavato*, che consiste nell'accoppiare ad ogni sillaba la nota corrispondente.

Da qui in poi, numerosissimi sono stati i compositori che hanno utilizzato questa tecnica, più genericamente definita *crittogramma musicale*, allargando l'associazione sillaba-nota alla notazione alfabetica. Il più celebre crittogramma è sicuramente B-A-C-H, utilizzato da Bach stesso in diverse occasioni, ad esempio come terzo soggetto dell'ultima fuga de L'arte della fuga, e successivamente da Schumann, Liszt, Maderna, Pärt e molti altri. Maurice Ravel utilizza il crittogramma H-A-Y-D-N<sup>2</sup> in Menuet sur le nom d'Haydn. Alban Berg, nel Kammerkonzert, ricava i soggetti del violino, del corno e del pianoforte dai nomi, in ordine, Anton Webern, Alban Berg e Arnold Schoenberg, mentre Pauline Oliveros in Dear John: a Canon on the name of Cage, utilizza le note Do, La, Sol e Mi dal crittogramma C-A-G-E<sup>3</sup>.

Dal soggetto cavato al crittogramma musicale, oggi è possibile trovare diverse applicazioni o siti web che permettono di sonificare un testo, ognuno seguendo i propri principi. *Typatone*<sup>4</sup> è un'applicazione che implementa l'associazione lettera-pitch basandosi su uno studio scientifico sulla ricorrenza di utilizzo delle singole lettere nella lingua inglese effettuato dalla Cornell University<sup>5</sup>. Vengono utilizzate le note naturali, mappate ai caratteri più frequenti nel range centrale e acuto, ai caratteri meno frequenti nel range basso e con gli accenti musicali prodotti dalle lettere meno frequenti in assoluto. *LangoRhythm*<sup>6</sup> è un sistema che associa le 26 lettere dell'alfabeto alle note naturali e ricava il ritmo analizzando la presenza delle vocali, costruisce accordi in base alla lunghezza della frase e sceglie la velocità di esecuzione in base alla lunghezza media delle parole<sup>7</sup>. Altri sistemi meno performanti che possono essere citati sono *P22 Music Text Composition Generator*<sup>8</sup> e *Melobytes*<sup>9</sup>.

Allo stato attuale, l'unica composizione pubblicata in anni recenti e documentata che utilizzi la sonificazione testuale attraverso il *mapping* carattere-pitch è "For Tho-

<sup>1</sup> **solmisazione** (o **solmizzazione**) s. f. [dal nome delle note musicali sol e mi]. In musica, teoria diffusa dal sec. XI a tutto il XVI e oltre, nella quale si limitava l'ambito tonale in una serie di 5 toni e d'un semitono, detta esacordo. [...] (da Treccani)

<sup>2</sup> Utilizzando la codifica in cui A e D corrispondono alle relative altezze, H al Si naturale (derivante dalla notazione tedesca), Y al Re naturale e N al sol naturale.

<sup>3</sup> <http://eamusic.dartmouth.edu/larry/dear.john/DearJohnNotes.html>

<sup>4</sup> <http://www.typatone.com>

<sup>5</sup> <http://works.jonobr1.com/Typatone>

<sup>6</sup> <https://kickthejetengine.com/langorhythm/>

<sup>7</sup> Informazione prese dalla conferenza TEDx "LangoRhythm, every speaker is a musician"  
<https://www.youtube.com/watch?v=N7a1Ua6kvcA>

<sup>8</sup> <https://p22.com/musicfont/>

<sup>9</sup> <https://melobytes.com/app/melobytes>

se *Who Died: A 9/11 Tribute*"[10]. Sono state effettuate ricerche attraverso il *web*, esplorando in particolare il sito *Sonifyer.org*<sup>10</sup> sul quale si trova un catalogo molto dettagliato di composizioni contemporanee che utilizzano la sonificazione, ma nessuna di questa implementa la sonificazione testuale.

Da un punto di vista scientifico, questo processo può essere riconosciuto nella definizione di *data sonification*: considerando i singoli caratteri di una frase come dati da codificare, "data sonification is the representation of data using sound"[4].

### 3. SONIFICAZIONE

La sonificazione è un sottotipo di *auditory display* che utilizza il suono non parlato per comunicare delle informazioni [3]. Utilizzata generalmente in contesti in cui la trasmissione delle informazioni non può avvenire su un piano visivo, in questo lavoro viene invece usata come strumento per forzare lo spostamento dell'attenzione dal contenuto della frase sonorizzata al processo di distorsione che rappresenta la generazione delle *fake news*.

La tecnica utilizzata è di tipo *Parameter-Mapping Sonification* [taxonomy] in cui i valori dei dati vengono mappati ad attributi timbrici e ritmici. Secondo una tassonomia degli *auditory display*, l'implementazione fatta rientra nella classe funzionale *data exploration*, in quanto la codifica della sonificazione ha come obiettivo quello di comunicare degli aspetti rilevanti del *data set* [3].

La codifica dei caratteri è di tipo *one to many*, in cui il dato di volta in volta codificato definisce sia il ritmo che il pitch. Il timbro viene invece generato con un tipo di codifica *many to one*, in cui sia la distanza della voce rispetto al centro dello spazio di diffusione sonora che la posizione nel tempo ne definiscono le caratteristiche spettrali [3]. La codifica e il processo di definizione ritmica sono ulteriormente approfonditi nei paragrafi 4.3 e 4.4.

### 4. IMPLEMENTAZIONE

Gli ambienti scelti per lo sviluppo di Distorsione di Massa sono il linguaggio di programmazione *SuperCollider*, installato su un computer *Raspberry Pi*, per la produzione e la gestione audio, messo in comunicazione via OSC con un eseguibile scritto in *Processing*, installato su un tablet con sistema operativo *Android*, utilizzato per la generazione della posizione delle voci nello spazio e la gestione dell'interazione con l'utente.

#### 4.1 Interfaccia

L'interfaccia del sistema è una postazione di lettura costituita da una sedia, un tavolo e da un tablet posizionato sul tavolo stesso, situata al centro dello spazio di diffusione sonora. Il tablet, agganciato in una posizione fissa, è appoggiato su un piano che nasconde uno spazio in cui sono collocati il *Raspberry Pi* e la scheda audio, connessi tra di loro via *usb*, alla quale sono cablati i quattro diffusori presenti ai quattro angoli della stanza. In questo spazio

<sup>10</sup> Forum per la ricerca sulla sonificazione

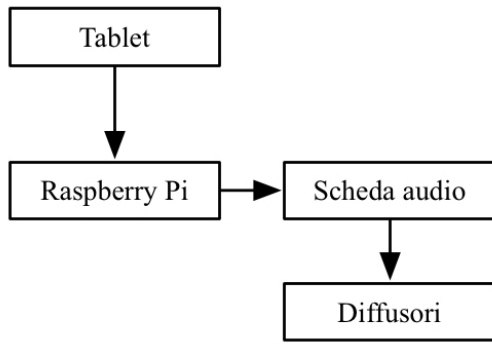


Figure 1. Flusso della comunicazione

vengono inoltre nascoste le prese elettriche per alimentare i dispositivi. La connessione tra il tablet e Raspberry Pi é di tipo wi-fi (Figura 1). L'interfaccia grafica é definita da un piano sul quale si muovono dei pallini di diverso colore che rappresentano la posizione delle voci nello spazio di diffusione. Sullo sfondo, un'immagine con frammenti di giornali permette di rafforzare la percezione del contesto di gioco (Figura 2).

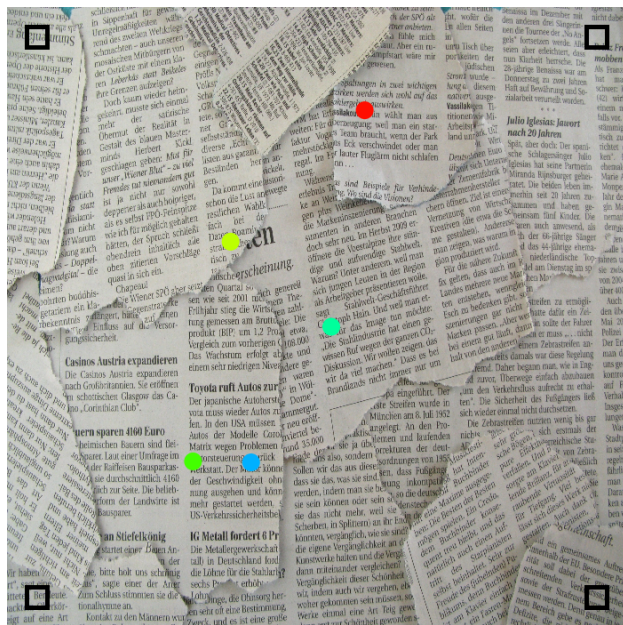


Figure 2. Interfaccia grafica

4.2 Interazione

L'utente che accede all'installazione trova il sistema in stato di attesa o con una sessione lasciata in sospeso da un utente precedente. Attraverso la pressione di un pulsante visualizzato sul tablet é possibile lanciare un comando di avvio/reset che fa partire il primo livello.

A questo punto viene riprodotto, sotto forma sonora, il titolo di una notizia ritenuta vera, posizionata al centro dello spazio. Dopo qualche secondo iniziano a comparire altre *n* voci (dove *n* corrisponde al numero del livello: una voce per il livello 1, due voci per il livello 2, e cosí via) che

riproducono all'unisono la stessa frase musicale. Superato qualche ulteriore istante, tutte le voci iniziano a muoversi nello spazio, ognuna con la propria traiettoria, e il processo di distorsione si avvia sulle *n* voci che nella fase iniziale hanno duplicato la prima.

Passato un lasso di tempo, la cui durata é direttamente proporzionale al livello di difficoltá, viene sbloccata la possibilitá di individuare e selezionare la voce rimasta inalterata attraverso la pressione diretta sul tablet. Se viene scelta quella giusta si passa al livello successivo oppure, in caso contrario, il livello riparte da capo.

4.3 Codifica Audio

Come punto di partenza, si é scelto di utilizzare come timbro una forma d'onda sinusoidale, come metafora della purezza della notizia vera da cui parte il processo. Si é successivamente deciso di definire una codifica melodico/ritmica per ogni singolo carattere che compone la frase.

Ogni carattere é codificato ritmicamente attraverso uno *step-sequencer* a cinque passi. Si é usato un sistema di codifica binario basato su 5 bit per poter facilmente codificare 32 caratteri (26 lettere e 6 caratteri di interpunzione). Per ognuno dei livelli é stata scelta una scala musicale at-



Figure 3. Esempio di codifica dei caratteri

traverso la quale vengono assegnati i pitch alle parole di bit che definiscono il ritmo. A partire da una frequenza fondamentale, assegnata al primo carattere nella lista delle codifiche, attraverso le relazioni tra i gradi della scala vengono assegnati i pitch ai successivi. Ogni carattere ha, quindi, una parola di 5 bit, che definisce il suo ritmo, e un pitch assegnato a una forma d'onda sinusoidale. (Figura 3, 4)

4.4 Processi di distorsione

Il processo metaforico di distorsione di una notizia, considerata vera allo stadio di partenza, viene realizzato parallelamente in tre modi:

- cambiando il contenuto della frase variando le singole parole con dei sinonimi<sup>11</sup> ;
- variando il contenuto spettrale del suono o l'articolazione dell'esecuzione;
- comprimendo o espandendo gli intervalli della scala scelta.

La variazione del contenuto della frase attraverso l'utilizzo di sinonimi viene effettuata accedendo a un file di

<sup>11</sup> A questo scopo é stato utilizzato il dizionario italiano dei sinonimi di OpenOffice, reperibile all'indirizzo <http://www.openoffice.org/it/linguistico/thesaurus.html#6>

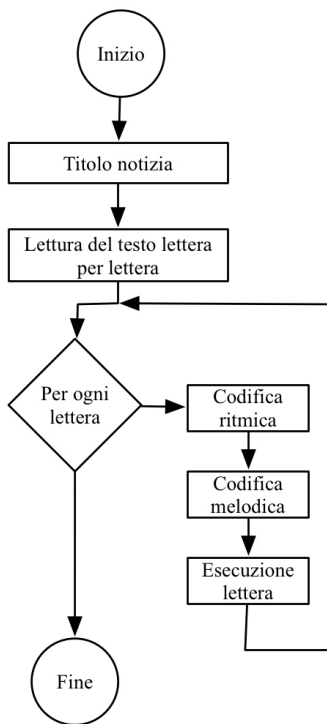


Figure 4. Diagramma codifica audio

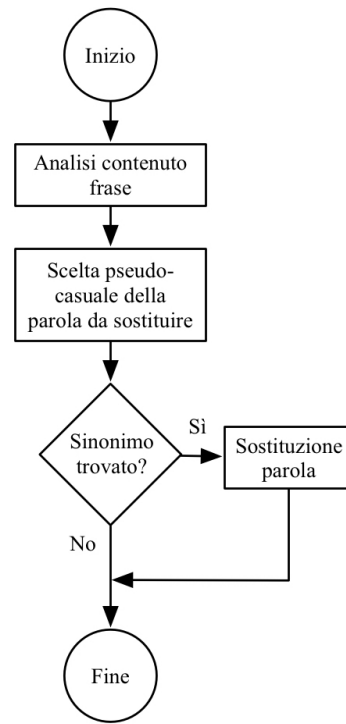


Figure 5. Diagramma sostituzione sinonimo

testo contenente un dizionario dei sinonimi. Alla fine della lettura del titolo da parte di ognuna delle voci, fatta eccezione per quella che rimane inalterata per tutta la durata dell'esecuzione, la frase viene analizzata e viene scelta in forma pseudo-casuale una parola da sostituire con un sinonimo. Se la ricerca di un sinonimo fallisce, vale a dire se la parola scelta non ha nessun sinonimo nel dizionario utilizzato, la frase rimane inalterata nella lettura successiva (Figura 5). La variazione del contenuto spettrale o dell'articolazione avviene attraverso l'utilizzo di *ring modulation*, *frequency modulation* e *wave shaping synthesis* per un arricchimento timbrico, *filtro comb/delay* per ottenere dalla variazione timbrica del phase shifting alla variazione dell'articolazione attraverso l'uso dinamico del *delay*, e *riverbero*, per le variazioni timbriche derivate dalle risonanze e dai differenti tempi di decay (Figura 6).

Il segnale di ognuna delle voci il cui contenuto varia nel corso dell'esecuzione, viene inviato a uno dei moduli sopracitati, i quali avranno un indice dinamico che varia sia 1) in relazione al tempo trascorso sia 2) all'istantanea posizione della voce nello spazio [8].

Sempre attraverso l'utilizzo di un indice dinamico relazione al tempo trascorso e alla distanza dal centro, può venire attuata una compressione/espansione degli intervalli della scala da cui vengono ricavati i singoli caratteri.

#### 4.5 Comunicazione

La comunicazione tra il dispositivo Android e il codice Supercollider su Raspberry Pi avviene attraverso il protocollo OSC e unidirezionalmente da Android a Supercollider.

I principali messaggi che vengono inviati sono:

- /start: avvia sia la parte grafica che la parte audio del gioco;
- /error: fa ripartire le sorgenti posizionandole al centro dello spazio;
- /posN (con "N" numero della sorgente): un array di due valori, normalizzati tra -1 e 1, che definisce la posizione nello spazio della sorgente.

#### 4.6 Spazializzazione

La spazializzazione viene gestita con l'utilizzo della tecnica di spazializzazione DBAP<sup>12</sup> [5] e l'implementazione del moto Browniano [8]. I dati relativi alla posizione vengono calcolati dall'applicazione compilata in Processing e successivamente inviati via OSC al motore audio in Supercollider su Raspberry Pi, che si occupa di posizionare le voci nello spazio sonoro.

Nonostante la posizione di interazione sia fissa (il tablet e la postazione su cui si trova sono al centro della stanza), il sistema DBAP da diversi vantaggi:

- nel caso non fosse possibile posizionare i quattro diffusori esattamente ai quattro angoli dello spazio di diffusione, la spazializzazione sarebbe ugualmente garantita;
- la potenza di calcolo necessaria alla computazione della posizione di una sorgente è molto inferiore rispetto ad altri tipi di spazializzazione: questo punto è cruciale in quanto la potenza di un Raspberry Pi rispetto a dei desktop o laptop è decisamente inferiore;

<sup>12</sup> Attraverso una classe sviluppata da uno degli autori: <https://github.com/StefanoCatena/dbapSC>

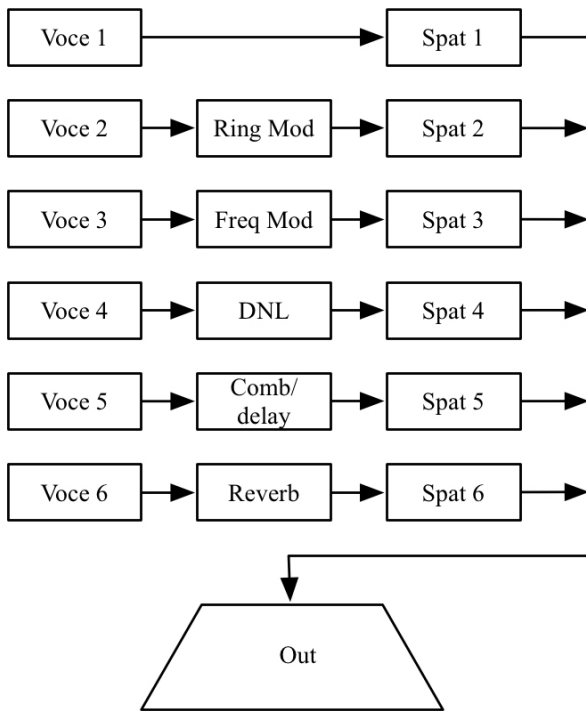


Figure 6. Diagramma processi audio

- i fruitori che hanno accesso allo spazio ma che non interagiscono direttamente con il sistema potrebbero muoversi mantenendo una buona percezione della posizione delle diverse sorgenti.

### 5. IL GIOCO

L'interazione tra l'utente e il sistema Distorsione di Massa avviene sotto forma di gioco: dalla postazione allestita al centro della stanza, si viene invitati ad ascoltare una traslazione sonora del titolo di una notizia vera. Ne viene messa in atto una moltiplicazione, una metaforica distorsione del contenuto e, infine, dopo un tempo prestabilito viene attivata la possibilità di riconoscere quale tra le voci che si stanno ascoltando è quella rimasta inalterata.

Nello stato di attesa, quando nessun utente sta interagendo con il sistema, viene diffuso un montaggio di voci di giornalisti che, in diverse lingue, annunciano delle notizie. In questo modo viene suggerito all'utente il contesto del gioco con l'obiettivo di farlo avvicinare al sistema e di farlo interagire.

Accanto al tablet si trova un foglio con il testo:

”La notizia nasce,  
la notizia si moltiplica,  
la notizia cambia forma.

Quando i simboli cambiano colore, trova la notizia rimasta inalterata.”

I livelli previsti sono cinque, ognuno con crescente difficoltà nel riconoscimento della voce inalterata rispetto al

precedente. La crescente difficoltà è direttamente proporzionale all'aumento di:

- range degli indici di distorsione;
- numero di distorsioni per voce;
- numero di voci;
- velocità di spostamento delle voci nello spazio;
- tempo necessario per l'attivazione delle possibilità di selezionare la voce inalterata.

Quando viene selezionata la notizia esatta, una scritta indica il risultato positivo, riportando anche il titolo della notizia, e si passa al livello successivo. Se la notizia selezionata è quella sbagliata, si viene informati dell'errore, viene riportata la notizia selezionata e il livello riparte da capo.

Superato l'ultimo livello viene visualizzato il messaggio "Ora continua a seguire la verità nella vita reale" e il sistema si spegne, in attesa di essere riattivato per ripartire dal primo livello.

### 6. LA COMPOSIZIONE

Se un utente decide di assistere senza interagire con il sistema, ciò che viene ascoltato è percepito come una vera e propria composizione di tipo algoritmico. Costituita da un motore centrale che converte una frase in una sequenza di ritmi associati a dei pitch, generando di fatto una partitura, evolve attraverso una strategia generativa basata su:

- la variazione del materiale testuale da traslare in forma sonora;
- l'elaborazione del suono in relazione al tempo;
- la posizione della voce nello spazio;

Come diffusamente esposto da Roads [6] nella definizione del quadro generale dei differenti approcci alla composizione algoritmica, anche Distorsione di Massa non può essere inquadrata in maniera esclusiva né come derivata da una tecnica deterministica né come derivata da una tecnica stocastica: le due dimensioni coesistono nel processo automatizzato di creazione sonora.

L'aspetto deterministico, vale a dire la parte del processo compositivo definita in modo tale che a un certo *seed data* corrisponda un solo risultato finale, si trova nella sonorizzazione di una frase: ogni carattere è mappato a una precisa codifica e, quindi, ogni volta che viene sottoposta al processo la stessa frase, il sistema produrrà la stessa frase musicale.

Gli aspetti stocastici, quelli definiti attraverso un processo di scelta di un dato secondo leggi probabilistiche, riguardano invece:

- la spazializzazione delle sorgenti, calcolate attraverso l'implementazione del moto Browniano;
- la variazione del contenuto lessicale, attraverso la scelta casuale tra i sinonimi delle singole parole;
- la distorsione sonora, definita dalla moltiplicazione tra l'indice generato in maniera deterministica dal passare del tempo e l'indice generato in maniera sto-

castica in relazione alla posizione della sorgente nello spazio.

La forma e il contenuto finali saranno definiti dalla capacità dell'utente che sta interagendo con il sistema di individuare la posizione della voce rimasta inalterata: la durata di ogni sezione, come anche le ripetizioni, dipendono dalla velocità con cui viene selezionata una delle voci e dall'esattezza della selezione.

## 7. PROBLEMI RISCONTRATI

Alcuni problemi si sono incontrati nel far comunicare l'applicazione compilata attraverso Processing su sistema Android e Supercollider in funzione su Raspberry Pi a causa dell'inaspettato malfunzionamento della libreria OSC P5 nella modalità Android di Processing: l'applicazione riesce a inviare messaggi a Supercollider ma non riesce a riceverli. L'imprevisto è stato risolto riscrivendo parte dei codici in modo che l'unica direzione di comunicazione sia quella dal dispositivo Android verso Raspberry Pi.

La gestione delle prestazioni di Raspberry Pi necessita di ulteriori test, in particolare per quanto riguarda la potenza di calcolo richiesta dal riverbero presente nella classe DBAP, utilizzato per posizionare le sorgenti nello spazio sonoro. Attraverso dei test, bisognerà giungere a un compromesso tra la completa implementazione delle riverberazioni richieste dalla tecnica in questione e l'effettiva resa del posizionamento delle voci anche con un minore utilizzo di elaborazioni del segnale ad alta richiesta di capacità di calcolo.

La codifica attraverso l'utilizzo di 26 caratteri è risultata efficace ma concettualmente limitata: ad esempio i diversi accenti e le maiuscole non vengono distinti, ma vengono convertiti nel loro corrispondente carattere neutro e minuscolo. Un futuro sviluppo potrebbe prevedere un maggior numero di caratteri codificati, variando anche la lunghezza delle parole di bit che definiscono il ritmo. Differenti lingue potrebbero avere differenti codifiche.

L'esecuzione del sistema attraverso l'utilizzo di Raspberry Pi genera un piccolo *lag* nella riproduzione sonora nel momento in cui viene attivata la funzione di ricerca di un sinonimo. Si è tuttavia scelto di non approfondire per il momento il problema in quanto si ritiene che questa componente dia anzi una variazione interessante nella resa sonora, piuttosto che una percezione di un errore.

## 8. CONCLUSIONI

Distorsione di Massa è stato testato in versione prototipale dagli autori. Si ritiene che il sistema abbia le potenzialità per raggiungere gli obiettivi posti in fase di progettazione, sia sul piano della rappresentazione metaforica della massa informe che risulta dalla distorsione delle informazioni che sul piano estetico. Ulteriori test verranno programmati ed effettuati prima di una esposizione pubblica in forma completa.

L'ideale sviluppo del sistema prevede l'inserimento della selezione dinamica della notizia da cui far partire il processo di distorsione. Il testo sarebbe ottenuto, in questo caso, accedendo a siti web di news. La difficoltà principale,

dato il tema trattato, consisterebbe nel testare la veridicità della notizia selezionata. Cambiando leggermente la natura del sistema, questo problema potrebbe essere risolto aggiungendo una fase in cui all'utente viene proposto un titolo preso dalla rete per il quale viene chiesto di giudicare se è vero o se è falso, quindi se può essere utilizzato per il gioco oppure no. Le risposte potrebbero essere raccolte e analizzate e i dati risultanti utilizzati in due modi: informare l'utente circa la coerenza della sua risposta rispetto alla risposta della massa e pilotare i segnali di controllo. Un'ulteriore espansione, seppur difficile da implementare, consisterebbe nel inserire un periodico *fact-checking* delle notizie analizzate, in forma automatizzata o umana, mettendo in relazione i dati con le risposte fornite dagli utenti per controllare il comportamento del sistema.

## 9. REFERENCES

- [1] H. Allcott, M. Gentzkow: "*Social Media and Fake News in the 2016 Election*", *Journal of Economic Perspectives*, 31(2), 211-36, 2017.
- [2] T. Hermann: "*Taxonomy and definitions for sonification and auditory display*", *Proceedings of the 14th Interational Conference on Auditory Display (ICAD 2008)*.
- [3] T. Hermann and A. Hunt: "*The Sonification Handbook*", (pp. 9-39, 363-397), J. G. Neuhoff Ed., Berlin, Germany: Logos Verlag, 2011
- [4] S. K. Lodha, J. Beahan, T. Heppe, A. J. Joseph, B. Zane-Ulman: "*MUSE: A Musical Data Sonification Toolkit*", *Proceedings ICAD 1997*.
- [5] T. Lossius, P. Baltazar, T. de la Hogue: "*DBAP Distance-Based Amplitude Panning*", *Proceedings ICMC 2009*, pp. 489-492.
- [6] C. Roads: "*Composing Electronic Music: a new aesthetic*", pp. 325-348, Oxford University Press, 2014.
- [7] C. Shao, G. L. Ciampaglia, O. Varol, A. Flammini, F. Menczer: "*The spread of fake news by social bots*", *arXiv preprint arXiv:1707.07592*, 2017.
- [8] A. Valle: "*Algorithmic Spatialization. Interfacing Supercollider to the Wave field system*", CIRMA/DAMS, Department of Humanities (StudiUm), University of Torino.
- [9] S. Vosoughi, D. Roy, S. Aral: "*The Spread of True and False News Online*", *Science*, Vol. 359, 2018.
- [10] M. Quinn, W. Quinn and B. Hatcher: "*For those who died: a 9/11 tribute*", Georgia Institute of Technology, 2003.

# GrainLab - Software open source per la sintesi granulare quasi-sincrona

Giovanni Sparano

Ricercatore indipendente

giovannisparano@gmail.com

## ABSTRACT

In questo articolo viene presentato *GrainLab*, software per la sintesi granulare in tempo reale. Nella sua implementazione si è prestata particolare attenzione alla scalabilità garantendo l'accesso al maggior numero di parametri, all'usabilità fornendo un'interfaccia grafica semplice e accessibile, e all'efficienza ottimizzando l'algoritmo di generazione dei grani. Il software è gratuito e *open source*, quindi facilmente modificabile e adattabile a esigenze specifiche.

## 1. INTRODUZIONE

La sintesi granulare nasce dalla teoria del quanto acustico, sviluppata da Dennis Gabor nel 1947 [1]. Il compositore Iannis Xenakis, nel suo *Formalized Music* [2], è stato il primo a concepire una teoria compositiva per i suoni granulari, mentre la prima implementazione digitale di un sintetizzatore granulare in tempo differito risale agli anni '70, grazie alle sperimentazioni di Curtis Road [3]. Nel 1986, Barry Truax realizza la prima implementazione di un software e la utilizza come unico strumento per la composizione del suo brano *Riverrun* [4]. A partire da questi primi esperimenti pionieristici, sono stati sviluppati moltissimi software per la sintesi granulare, ognuno caratterizzato da un approccio alla sintesi differente in base a necessità sia compositive sia tecnologiche.

*GrainLab* è un granulatore quasi-sincrono[5] il cui principio di funzionamento si basa su linee di ritardo opportunamente finestrate. Il suo obiettivo è quello di fornire uno strumento duttile, non oneroso dal punto di vista delle risorse e facilmente adattabile alle necessità più disparate. È gratuito e *open source* ed è prevista un'ampia gamma di parametri di controllo: numero di grani, intervallo di ampiezze e densità probabilistica dei grani, fase, ritardo del segnale in ingresso, trasposizione, finestatura e panning per singolo grano.

Un prima versione del software è stata utilizzata per la realizzazione del *live electronics* dell'opera multimediale *FENIX DNA* di Fabrizio Plessi, su musiche dell'autore, andata in scena al Teatro La Fenice di Venezia dal 25 luglio al 6 agosto 2017<sup>1</sup>. Il software prevedeva l'utilizzo di

<sup>1</sup> Maggiori informazioni si possono reperire all'indirizzo web <http://www.teatrolafenice.it/site/index.php?pag=21&spettacolo=25026>

Copyright: ©2018 Giovanni Sparano et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution License 3.0 Unported](https://creativecommons.org/licenses/by/3.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

5 istanze di *GrainLab*, una per ogni strumento coinvolto nell'esecuzione:

- flauto;
- clarinetto basso;
- viola;
- pianoforte;
- voce di soprano,

di 4 delay spettrali, uno per ogni strumento acustico a eccezione della voce, e di un sistema di spazializzazione del suono per un sistema di diffusione multicanale.

## 2. IMPLEMENTAZIONE

L'implementazione di *GrainLab* è stata realizzata grazie all'ausilio di *Max/MSP*, linguaggio di programmazione visuale sviluppato da Cycling '74 [6], sfruttando in particolare le potenzialità offerte da *Gen*, introdotto nel 2011 con la versione 6. Attraverso quest'oggetto è possibile definire processi di elaborazione del segnale audio con un'accuratezza al campione e un dispendio di risorse estremamente limitato. Tutti i *patcher* in *Gen* vengono compilati al momento del salvataggio, permettendo un livello di astrazione più basso rispetto agli algoritmi implementati in *MSP*. Il software è reperibile presso il sito internet dell'autore<sup>2</sup>.

### 2.1 Schema generale

Il generatore di grani che sta alla base di *GrainLab*, riceve in ingresso un segnale audio monofonico, un segnale rampa di sincronizzazione globale, una serie di controlli per la gestione dei parametri di granulazione, e restituisce in uscita un segnale stereofonico (vedi fig. 1).

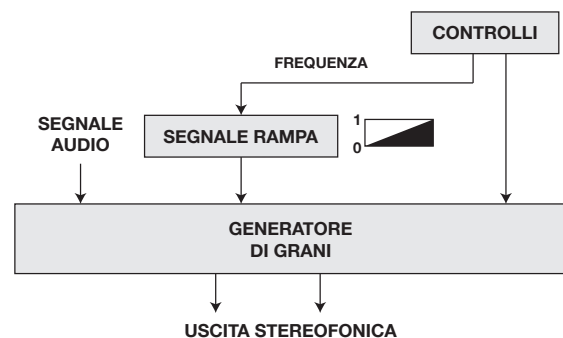


Figura 1. Schema generale.

<sup>2</sup> <http://www.giovannisparano.it/attivita/maxmsp.php>

### 2.2 Polifonia e sincronizzazione

Il numero di grani polifonici è modificabile tramite un apposito controllo dell'interfaccia grafica e varia tra 1 e  $N$ , con  $N$  dipendente dalle risorse di calcolo dell'elaboratore.

Tutti i grani sono pilotati da un singolo segnale rampa di sincronizzazione, con frequenza inversamente proporzionale alla lunghezza dei grani. Per ogni istanza di grano è prevista la possibilità di sommare al segnale rampa di sincronizzazione un numero reale, variabile tra 0 e 1 e impostabile tramite tabella. Il segnale rampa viene poi opportunamente limitato nell'intervallo tra 0 e 1 grazie a una funzione *modulo* che restituisce il resto della divisione euclidea in numeri reali tra 0 e 1.

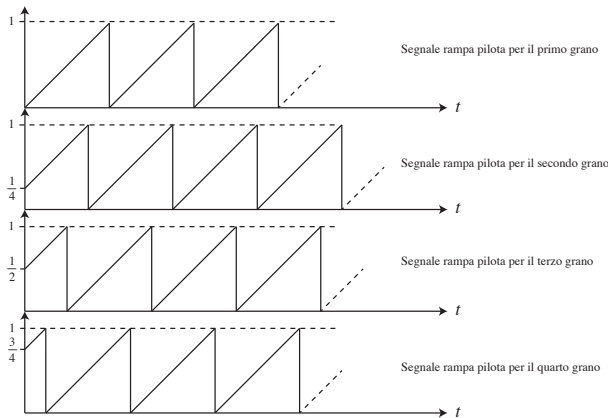


Figura 2. Esempio di sfasamento del segnale rampa di sincronizzazione con 4 grani.

Vengono forniti due preset, *continuous* e *rhythmic*, per la gestione della distribuzione temporale dei singoli grani. Per il primo preset, i grani sono distanziati tra loro di un valore di fase pari all'inverso del numero di grani. Ad esempio, con 4 grani, il primo grano sarà sincrono al segnale rampa di sincronizzazione, il secondo dopo  $\frac{1}{4}$  di fase, il terzo dopo  $\frac{1}{2}$ , e il quarto dopo  $\frac{3}{4}$  (vedi fig. 2). Per il secondo preset, metà dei grani sono sincroni senza sfasamento e l'altra metà viene sfasata di  $\frac{1}{2}$  di fase.

### 2.3 Lettura e finestatura

Ogni grano riceve in ingresso un segnale audio monofonico e lo ritarda di un valore casuale compreso nell'intervallo impostato dall'utente. I valori accettati variano tra 0 e 30 secondi, con una risoluzione al millisecondo.

Le funzioni di trasferimento della finestatura possibili sono 6, in 9 differenti preset. La finestatura avviene grazie alla lettura di un buffer di 512 campioni sincronizzato con il segnale rampa.

Il primo preset è la finestra *Hann* con funzione:

$$w_n = \frac{1}{2} \left[ 1 - \cos \left( \frac{2\pi n}{N-1} \right) \right], \quad (1)$$

con  $N$  uguale a 512 e  $n$  da 0 a  $N-1$ .

Il secondo preset è la finestra *Expodec*[7] con funzione:

$$w_n = e^{\frac{-2\alpha|n-N|}{N-1}}, \quad (2)$$

con  $\alpha$  uguale a 3,65, mentre il terzo preset è la finestra *Rexpodec*, con funzione uguale alla precedente *Expodec* ma con valori di  $n$  invertiti, ossia da  $N-1$  a 0.

Il quarto preset è la finestra *Triangle*, ossia una rampa lineare ascendente da 0 a 1 per la prima metà del buffer e discendente per la seconda metà.

Il quinto preset è la finestra *Trapezoid*, ossia una rampa lineare ascendente per un terzo del buffer, 1 per un altro terzo e una rampa discendente per la restante porzione.

Il sesto preset è la finestra *Sinc* con funzione:

$$w_n = \frac{\sin n}{n}, \quad (3)$$

con  $n$  tra  $-4\pi$  e  $4\pi$ , ossia con 4 passaggi a 0.

Il settimo preset include tutte le possibili finestre viste in precedenza, assegnate a gruppi di 6 grani successivi.

L'ottavo e il nono preset prevedono rispettivamente l'alternanza delle finestre *Expodec* e *Rexpodec* e di quelle *Hann* e *Sinc*.

### 2.4 Densità

La densità dei grani è regolata da un algoritmo booleano che genera un *duty cycle* del segnale rampa di sincronizzazione (vedi fig. 3).

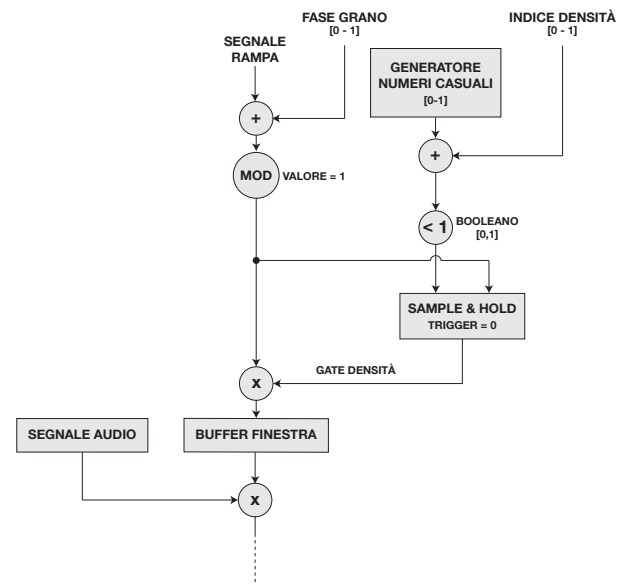


Figura 3. Algoritmo per la gestione della densità.

Viene sommato un valore di densità variabile tra 0 e 1 a un numero casuale, generato a frequenza di campione, compreso tra 0 e 1. Per valori di questa somma minori di 1, l'operatore booleano restituisce 1, altrimenti 0. Un *Sample and Hold* campiona questo risultato e accetta come trigger il segnale rampa a fase 0. L'uscita del *Sample and Hold* viene moltiplicata per il segnale rampa restituendo, nel caso in cui l'operatore booleano sia falso, un valore pari a 0 per un intero periodo di rampa. Il segnale audio in ingresso verrà di conseguenza moltiplicato esclusivamente per il primo campione del buffer di finestatura, opportunamente impostato a 0, generando un segnale audio in uscita di ampiezza nulla.

Il parametro di densità impostato dall'utente, espresso in valore percentuale, aumenta o diminuisce la probabilità che il *duty cycle* occorra.

### 2.5 Trasposizione, panning e ampiezze

Vengono fornite diverse tabelle preimpostate per le trasposizioni e per il panning per singoli grani. L'intervallo utilizzabile per le trasposizioni varia tra un'ottava sotto e un'ottava sopra la sorgente elaborata. Sia per le trasposizioni, sia per il panning, è possibile interpolare i diversi preset in un tempo definibile dall'utente.

La variazione dell'ampiezza ha un principio di funzionamento analogo a quello dei ritardi, ossia l'assegnazione casuale di un valore all'interno di un intervallo espresso in *decibel*.

### 2.6 Gestione dei parametri di controllo

Tutti questi parametri di controllo, ad eccezione del numero di grani, possono essere cambiati in fase di funzionamento del software senza generare artefatti, click o interruzioni del segnale elaborato in uscita. I cambi di parametro avvengono a fase 0 del segnale rampa, attraverso un *Sample and Hold*. In quell'istante la finestatura, sincronizzata anch'essa con la rampa, restituisce un'ampiezza nulla di segnale audio in uscita (vedi fig. 4).

Gli immediati sviluppi futuri del software da parte dell'autore saranno rivolti all'aggiunta di un sistema di uscite multicanale, all'ampliamento delle funzioni di trasferimento per la finestatura e a fornire la possibilità di elaborare una sorgente preregistrata.

## 4. BIBLIOGRAFIA

- [1] D. Gabor, "Acoustical quanta and the theory of hearing," *Nature*, vol. 159, pp. 591-594, Maggio 1947.
- [2] I. Xenakis, *Formalized Music*. Bloomington: Indiana University Press, 1971.
- [3] C. Roads, "Introduction to granular synthesis," *Computer Music Journal*, vol. 12, no. 2, pp. 11-13, 1988.
- [4] B. Truax, "Real-time granular synthesis with a digital signal processor," *Computer Music Journal*, vol. 12, no. 2, pp. 14-26, 1988.
- [5] C. Roads, *The Computer Music Tutorial*. Cambridge, MA, USA: The MIT Press, 1996.
- [6] "https://cycling74.com/."
- [7] C. Roads, *Microsound*. Cambridge, MA, USA: The MIT Press, 2004.

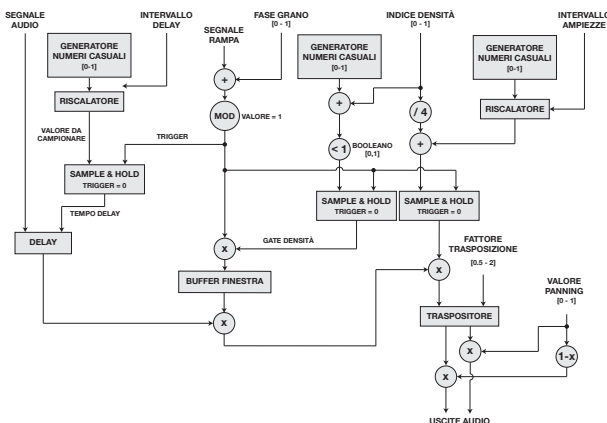


Figura 4. Schema di un'istanza di grano.

Relativamente ai parametri di fase e finestatura, è previsto un procedimento differente: viene imposto al singolo grano (in successione e con un ritardo progressivo) un *duty cycle* del segnale di sincronizzazione, si effettua la reinizializzazione e infine si libera il segnale rampa nell'istante di fase 0 successivo. La durata di questi cambi di preset è definibile dall'utente (valore minimo  $500N$  millisecondi, con  $N$  uguale a numero di grani).

## 3. CONCLUSIONI E SVILUPPI FUTURI

*GrainLab*, data la sua natura open source, si presta a ulteriori sviluppi, sia da parte dell'autore, sia di un qualsiasi utilizzatore che necessiti di particolari configurazioni o personalizzazioni del codice.

# Electroacoustic and Computer Music Timelines

**Paolo Zavagna**

Conservatorio “B. Marcello”  
Venezia  
paolo@zavagna.it

**Francesco Peccolo**

Conservatorio “B. Marcello”  
Venezia  
francescopeccolo96@gmail.com

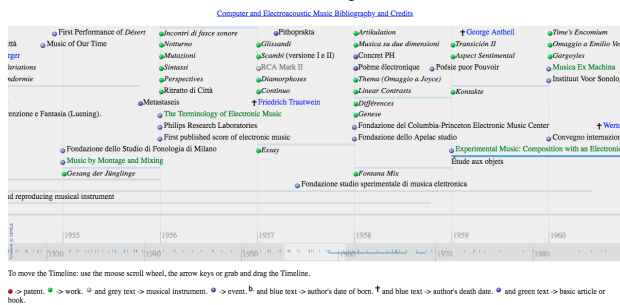
**Riccardo Sellan**

Conservatorio “B. Marcello”  
Venezia  
riccardo.sellan@hotmail.it

## ABSTRACT

Electroacoustic [4] and Computer Music [3] timelines are projects created by Paolo Zavagna in collaboration with students from Conservatories of Venezia and Padova (Italy). Timelines are organized by events, biographies, compositions, technologies and other relevant topics about electroacoustic and computer music, each one viewable with additional information. The project is totally open-source and uses the SIMILE widget (MIT) [1] that adapts to the most commonly-used browsers. In the figure below you can see an overall view of one of the two websites.

Electroacoustic Music Timeline / Cronologia della musica elettroacustica



## 1. INTRODUCTION

The purpose is to create a tool for researchers, students and anyone who is interested in electronic and computer music history and research; another purpose is to popularize the field of electronic and computer music. Timelines include historical events since the end of the 19<sup>th</sup> century. Electroacoustic Music timeline includes a period approximately before the 1970's while Computer Music covers the following years. Information is divided into two different timelines for better consultation.

## 2. FIRST LOOK

Time flux is disposed horizontally with the latest year in the right side. Under the main timeline there is a scroll menu with a larger time window that shows the density of events through several decades; the main time window is about eight years (it depends on the monitor resolution).

Copyright: ©2018 Paolo Zavagna et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution License 3.0 Unported](https://creativecommons.org/licenses/by/3.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Events are distinguished by colors and symbols described in a key at the bottom of the page (see figure below).



## 3. SOURCES OF INFORMATION

All sources of information are listed in a bibliographic file – periodically updated and both in pdf and bib format – that you can download from the websites. As you can see in the following example, each record, in json format [2], has a field for the “source” of the information.

```
{
  'start': '1996TZ',
  'title': '<i>Me-Oh</i>',
  'description': "by Shlomo Dubnov, composition for Voice, ensemble and live electronics, commissioned by the Tarshish Ensemble, performed at the Jerusalem Festival, 1996.",
  'caption': "Shlomo Dubnov",
  'icon': "green-circle.png",
  'source': "G. Fronzi, Electrosound, 2013, p.152"},
}
```

## 4. APPROACH

### 4.1 Choosing timelines

Timelines include historical events since the end of the 19<sup>th</sup> century. Electroacoustic Music timeline includes a period approximately before the 1970's and events and people more involved in electroacoustic ‘analog’ music production while Computer Music covers the years starting from the late 1950's and musical production by means of digital or mixed analog plus digital systems.

### 4.2 Moving in the timeline

There are several ways to consult timelines. The first method is to drag the main bar to the left or right for a fine shift; on the other hand it is possible to do the same action to the menu for a larger shift.

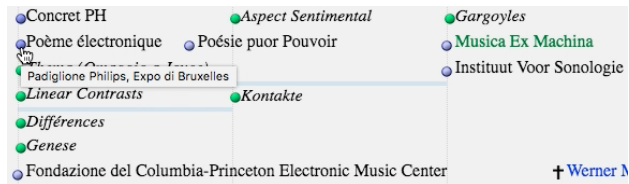
The drag method moving through the years can also be achieved by double-clicking on the left or on the right side of timelines. A double-click on the main timeline will produce a smaller movement while double-clicking on the menu bar causes a larger movement.

Another method is to scroll the mouse wheel to do a precise shift; changing direction of the scroll can be used to move through different time directions.

The last method is to use the arrow keys, up and left make the older events viewable while the down and right show earlier ones.

### 4.3 Scrolling mouse

By scrolling the mouse on a topic, basic information can be read about it (place and city of an event, for example), as you can see in the figure below.



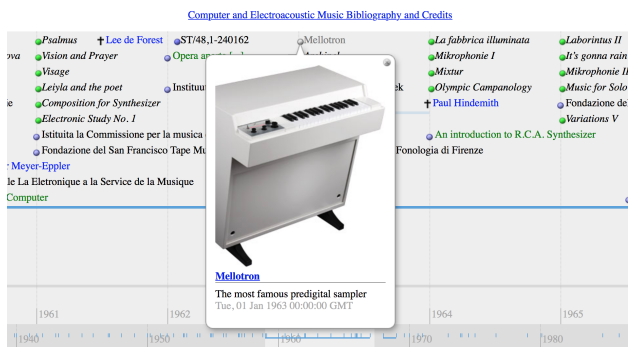
### 4.4 Select an event

To select an event it's enough to click on the symbol or on the linked label; doing this operation, a pop-up window will appear. This window shows some essential information.

### 4.5 Reading information

The pop-up window of each event shows brief descriptions, images and precise date if it's known (see figure below). Most of them are linked with external websites that provide more specific details or multimedia contents.

Electroacoustic Music Timeline / Cronologia della musica elettroacustica



## 5. KEY

The key is located at the bottom of the web page and illustrates the different kind of topics.

### 5.1 People

Brief biographies regarding composers, engineers, inventors and others are marked with blue text for the name.

#### 5.1.1 Birth

The date of birth of a composer and relevant personalities is indicated by the symbol **b.** to the left of the linked name.

#### 5.1.2 Death

The date of death of a composer and relevant personalities is indicated by the symbol **†** to the left of the linked name.

### 5.2 Compositions and works

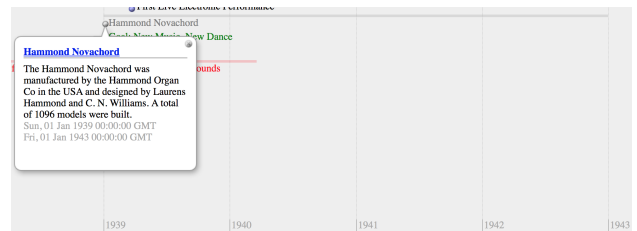
Compositions and other relevant musical works are indicated with a green dot to the left of the linked work.

### 5.3 Articles or books

Basic articles and books on the field of electroacoustic and computer music are indicated by green text and a blue dot to the left of the linked label.

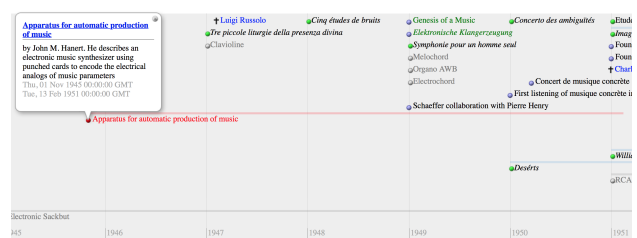
### 5.4 Musical instruments

Electronic, electroacoustic and digital instruments are indicated by grey text and a dot to the left of the linked instrument. The 'duration' of the musical instrument, marked by a grey line, is related to its presence in the marketplace (see for example the Hammond Novachord, produced between 1939 and 1942, in the figure below).



### 5.5 Patents

Patents are marked by a red dot to the left and red text. The 'duration' of the patent, marked by a red line, starts when the patent has been filed and ends when it was patented.



### 5.6 Events

Relevant events like concerts, meetings and foundations are indicated by a blue dot to the left of the linked event. The duration of the event is marked by a light blue line.

## 6. CONTRIBUTE

### 6.1 Creative Commons

The whole project is licensed under the terms of the Creative Commons license with some rights reserved: the ownership has to be always recognized, timelines can be used without commercial intents, the resulting materials linked to the project must have the same license.

## 6.2 Active participation

Electroacoustic and Computer Music timelines are created with the SIMILE timeline widget developed and powered by MIT [1]. Timelines can be edited with the JSON format [2] to create and modify events and other structures. The work is in progress and people are invited to report misprints, errors, omissions to the authors.

## 6.3 Future updates

The possible future updates and improvements will include a search engine to make consultation easier with keywords and tag. Also, the timeline time window can be changed. Anyone can access the open-source code and propose improvements.

## 7. REFERENCES

- [1] David François Huynh. *Timeline. Web Widget for Visualizing Temporal Data*. URL: <http://www.simile-widgets.org/timeline/>.
- [2] *Introducing JSON*. URL: <https://www.json.org>.
- [3] Paolo Zavagna, ed. *Computer Music Timeline*. URL: <http://www.computermusictimeline.eu/>.
- [4] Paolo Zavagna, ed. *Electroacoustic Music Timeline*. URL: <http://www.electroacousticmusictimeline.eu/>.

# InMuSIC: un sistema interattivo multimodale per l'improvvisazione elettroacustica

Giacomo Lepri

Centre for digital music - Queen Mary University of London

g.lepri@qmul.ac.uk

## ABSTRACT

InMuSIC è un sistema interattivo concepito per l'improvvisazione elettroacustica (clarinetto e live electronics). Attraverso un'analisi multimodale del comportamento del musicista (produzione del suono e movimento del corpo) InMuSIC genera e organizza nuovi materiali sonori. Gli interventi audio originati rappresentano una integrazione musicale in dialogo con le sonorità articolate dallo strumentista. La codifica e interpretazione di un complesso insieme di interazioni musicali e motorie permettono quindi al sistema di bilanciare differenti gradi di adattività e autonomia musicale.

## 1. INTRODUZIONE

Nel contesto della musica occidentale l'improvvisazione ha spesso avuto un ruolo rilevante. Le prassi vocali del Gregoriano e dell'Ars antiqua, la cadenza libera e le sperimentazioni relative alla musica contemporanea e al free jazz del Novecento sono solo alcuni degli esempi che permettono di considerare l'improvvisazione come una componente autorevole di molte tradizioni musicali europee. A partire dagli anni Sessanta del XX secolo compositori, strumentisti e performers hanno iniziato ad esplorare la pratica dell'improvvisazione anche all'interno dei contesti relativi alla musica elettronica. Il tentativo di sviluppare convincenti e spontanee interazioni musicali attraverso l'uso di nuove tecnologie e di nuovi linguaggi musicali sta tuttora spostando (se non dissolvendo) i confini tra composizione ed improvvisazione nella cultura musicale occidentale.

Con l'avvento delle tecnologie digitali si è sviluppato all'interno del Sound and Music Computing un nuovo campo di studi centrato sul design di sistemi interattivi concepiti per essere utilizzati in contesti improvvisativi. Queste ricerche di carattere compositivo, vengono spesso supportate da discipline extramusicali sia scientifiche che umanistiche (e.g. psicologia cognitiva, neuro scienze e human-computer interaction). I sistemi interattivi musicali offrono infatti uno scenario ideale per condurre studi relativi alla percezione, alla comunicazione/espressione, ai processi creativi e allo sviluppo di nuove tecnologie.

Copyright: ©2018 Giacomo Lepri et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution License 3.0 Unported](https://creativecommons.org/licenses/by/3.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

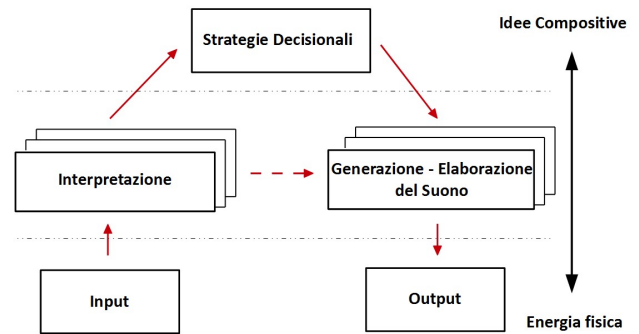


Figure 1. Il modello concettuale del sistema interattivo InMuSIC.

Negli ultimi vent'anni l'interesse al design di sistemi interattivi in grado di dialogare con i musicisti è significativamente cresciuto. I sistemi progettati dal compositore/improvvisatore George Lewis *Voyager* [1] e dal ricercatore informatico François Pachet *Continuator* [2] rappresentano due delle applicazioni musicali più significative recentemente sviluppate.

Gradualmente le esperienze acquisite da compositori, interpreti e ricercatori grazie alle indagini multidisciplinari inerenti alla composizione algoritmica [3] e alla liuteria digitale [4] forniscono una sempre più proficua cognizione del tema. In questo contesto sono evidenti i rapporti tra le strategie di progettazione e realizzazione di una specifica tecnologia/software ed il paradigma compositivo definito per sviluppare le interazioni musicali. Il framework utilizzato per la realizzazione di InMuSIC riflette quindi particolari intuizioni compositive e strategie improvvisative maturate all'interno di questo specifico contesto culturale.

## 2. IL MODELLO COMPOSITIVO

La definizione di un modello concettuale di riferimento rappresenta uno degli aspetti fondamentali nello sviluppo di sistemi interattivi musicali. L'esplicitazione del modello teorico permette infatti sia di chiarire la struttura implementativa nel suo insieme sia di organizzare le singole unità del sistema a seconda delle loro specifiche funzioni. Comporre interazioni musicali convincenti ed efficaci è una delle principali sfide che questo tipo di ricerca pone. Adottare un solido e ponderato modello compositivo agevola l'adempimento di questo complesso onere e facilita le varie fasi implementative. L'atto compositivo deve quindi tenere in considerazione ogni singola fase del processo interattivo (dalla cattura ed organizzazione delle informa-

zioni in ingresso fino alla articolazione dei materiali sonori generati durante l'improvvisazione).

In questa sezione viene presentato il modello concettuale di riferimento per la realizzazione di InMuSIC. La struttura del sistema si basa sul modello proposto da Leman e Camurri [5] per lo sviluppo di piattaforme interattive multimodali rivolte all'analisi dell'espressività musicale. La struttura è composta da cinque moduli collocati su tre piani concettuali (figura 1). Ciascun piano è associato ad un differente grado di astrazione. È possibile quindi tracciare un continuum che partendo dalla rappresentazione del mondo fisico si estende verso una dimensione musicale associata ad idee ed intuizioni compositive.

Il livello più basso è quindi attinente a quelle unità del sistema che svolgono compiti connessi al dominio fisico (e.g. rilevamento dei movimenti e delle sonorità prodotte dal musicista). Il livello più alto è invece relativo alle componenti più astratte del sistema, responsabili delle scelte compositive che concretizzano le interazioni musicali in tempo reale. Le funzioni dei singoli moduli sono quindi associate al trattamento dei dati in ingresso (Input e Interpretazione), in uscita (Generazione - Elaborazione del Suono) e alle dinamiche che definiscono le relazioni tra i vari processi di input ed output. Questa rappresentazione esplicita un ciclo interattivo distribuito su diversi piani concettuali e offre la possibilità di comporre e sviluppare quello che sarà il comportamento musicale del sistema.

Il modello presentato potrebbe essere interpretato come una possibile strategia di mapping, tuttavia, in questo contesto, si preferisce non fare strettamente riferimento al termine mapping. Esso sembra infatti perdere efficacia in quelle situazioni in cui processi algoritmici complessi vengono implementati per influenzare il comportamento del sistema e quindi la generazione del materiale sonoro [6]. Le funzioni dei singoli moduli e le loro relazioni vengono, di seguito, brevemente descritte.

- *Input*: il modulo si distingue per due funzioni principali: rilevare i movimenti e le sonorità prodotte dal musicista e convertire questa energia (cinetica e sonora) in informazione digitale. InMuSIC prevede l'utilizzo di due sensori: il suono dello strumento viene captato usando un microfono a condensatore, il movimento del performer viene catturato tramite un sensore di movimento 3D (Microsoft Kinect 2).
- *Interpretazione*: all'interno di questa unità le informazioni audio e gestuali in ingresso vengono analizzate ed interpretate attraverso diversi processi paralleli. In questa fase vengono estratte specifiche informazioni di carattere musicale ed espressivo significative del comportamento del musicista. Durante questa analisi la quantità di informazione diminuisce notevolmente poichè questo modulo filtra dati specifici di carattere prettamente musicale ed espressivo.
- *Strategie decisionali*: il modulo è collocato al livello più alto di astrazione musicale all'interno del modello compositivo. La sua funzione principale è relativa alla organizzazione temporale dei processi

adibiti alla generazione e manipolazione del suono. Le strategie decisionali implementate si basano principalmente sulla continua correlazione delle analisi di suono e movimento contestualizzate in differenti finestre temporali.

- *Generazione*: questa unità è composta da diversi algoritmi adibiti all'elaborazione digitale del suono: i materiali sonori proposti dal sistema vengono qui concretamente generati ed elaborati. Il sistema può assegnare, al fine di stabilire interazioni dirette con ciò che il musicista propone, il controllo di alcuni dei parametri degli algoritmi per la generazione del suono ai dati estratti in tempo reale dai moduli relativi all'analisi del comportamento sonoro e motorio dell'esecutore.
- *Output*: il modulo è adibito al trasferimento nel dominio fisico delle informazioni sonore generate dalle unità più astratte. I processi principali coinvolti in quest'ultima fase sono relativi all'amplificazione e compressione del segnale generato, alla conversione del segnale da digitale ad analogico e alla proiezione del suono nello spazio performativo.

### 3. ARCHITETTURA DEL SISTEMA

Da un punto di vista implementativo è possibile suddividere l'architettura del sistema in quattro unità principali: analisi del movimento, analisi del suono, strategie decisionali e generazione del suono (figura 2). Ogni unità è assegnata a componenti software specifiche comunicanti tramite protocollo Open Sound Control (OSC). Il sistema è quindi caratterizzato da un approccio modulare ed è possibile eseguire le varie componenti su diversi calcolatori connessi tra loro. Una descrizione dei quattro moduli e delle loro funzioni viene presentata di seguito.

#### 3.1 Analisi del suono

L'unità cattura e filtra le informazioni provenienti da un microfono condensatore a clip. Differenti tecniche per l'estrazione in tempo reale di specifiche qualità sonore vengono eseguite parallelamente (Max/MSP). Nello specifico questo modulo compie analisi relative al rilevamento dell'ampiezza (loudness), attacco (onset) [7], frequenza fondamentale (pitch variation) [8], rumorosità (noisiness) [9] e baricentro spettrale [7] (spectral centroid).

Durante l'improvvisazione le informazioni audio vengono costantemente interpretate e ridefinite. Tramite una costante comparazione dei dati forniti dalle varie analisi sonore il sistema sceglie l'informazione più attendibile (da un punto di vista percettivo). Un semplice esempio di questo procedimento può essere la strategia utilizzata per effettuare il rilevamento dell'attacco: in questa unità sono implementati infatti tre algoritmi per l'estrazione di informazioni relative all'attacco sonoro. Il modulo seleziona una delle tre informazioni disponibili in relazione alle informazioni provenienti dall'analisi dell'ampiezza dello stesso

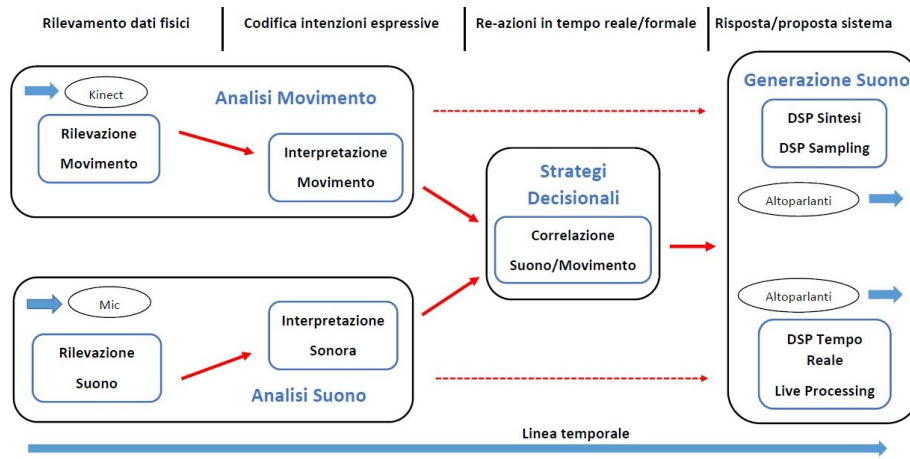


Figure 2. Il flusso dati nella struttura implementativa di InMuSIC.

segnale audio (ogni singolo algoritmo per l'estrazione degli attacchi è più affidabile in relazione a determinati range dinamici).

L'abilità del sistema di scegliere differenti strategie per l'analisi del suono consente quindi una rilevazione delle informazioni musicali di carattere adattivo. La metodologia utilizzata per l'interpretazione dei materiali sonori prodotti dal musicista si ispira alla straordinaria capacità umana di discriminare e selezionare specifiche informazioni musicali all'interno di contesti sonori caratterizzati da un alto grado di complessità. Il tentativo è quindi quello di modellare, con una accettabile approssimazione, questa abilità.

### 3.2 Analisi del movimento

L'unità esegue l'analisi del movimento per mezzo di vari algoritmi per l'estrazione in tempo reale di specifici gesti espressivi full body [10]. I dati provenienti dal sensore 3D vengono elaborati tramite la piattaforma EyesWeb [11]. Sulla base della definizione di gesto fornita da Kurtenbach e Hultheens [12]: "un movimento del corpo che contiene informazione", il gesto può essere definito espressivo se questa informazione è caratterizzata anche da contenuti di natura espressiva, cioè veicola potenziali messaggi impliciti [13]. Spesso queste informazioni aggiuntive introducono all'interno del processo comunicativo nuove qualità interpretabili, le quali solitamente, pur godendo di una loro autonomia espressiva, vengono giustapposte a specifici significati denotativi.

Sulla base di queste considerazioni, l'approccio adottato utilizza il modello sviluppato da Camurri et al. per l'analisi di gesti espressivi tramite l'impiego di sistemi interattivi multimodali [14]. La nozione di movimento espressivo presa in considerazione implica caratteristiche motorie associate ad un'ampia gamma gestuale non ristretta a tipologie di movimento specifiche. Il modulo non è quindi concepito per riconoscere gesti particolari, la ricerca è invece indirizzata ad una esplorazione del movimento di carattere qualitativo. Certamente ogni strumento (tradizionale o relativo alle nuove tecnologie) è connotato da un specifico apparato di affordances gestuali [15]. Per di più, ogni

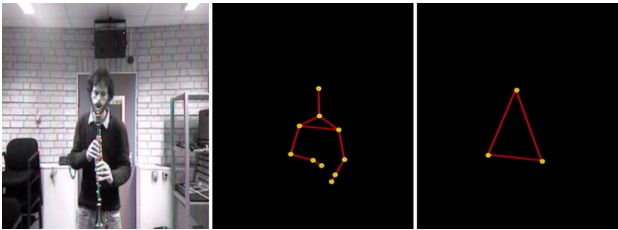
musicista sviluppa una personale modalità di interazione fisica con esso.

La sfida consiste proprio nell'implementare algoritmi per l'estrazione di informazioni di carattere astratto e generalizzato, rappresentative di ampie sfere comunicative associate al movimento espressivo. La strategia adottata prevede quindi una semplificazione (minima rappresentazione) del movimento espressivo attraverso un'analisi di tipo qualitativo applicata alla parte superiore del corpo del musicista [15]. Considerando un numero ridotto di informazioni visive (Fig. 4) (posizione, velocità e accelerazione di testa, mani, spalle e gomiti del musicista), vengono effettuate tre analisi del movimento di carattere espressivo: fluidità del movimento (fluidità e continuità vs impulsività), variazione della postura (chiusa vs aperta) ed energia cinetica (variazione della quantità di movimento della parte superiore del corpo) [15].

### 3.3 Comparazione suono-movimento

In linea con l'approccio di rappresentazione minimale precedentemente introdotto, le informazioni provenienti dagli algoritmi di analisi vengono interpretate in relazione a tre possibili stati: *high*, *low* o *stable*. Prendendo ad esempio la feature audio, è possibile descrivere e associare i vari stati ai seguenti tratti musicali: (i) dinamica *high*, *low* o *stable* (crescendo vs. diminuendo); (ii) *high*, *low* o *stable* rilevazione di attacchi (aumento o diminuzione della densità degli eventi); (iii) *high*, *low* o *stable* deviazione della frequenza fondamentale (espansione vs. riduzione della gamma di frequenza utilizzata). Questa classificazione viene effettuata per tutte le features analizzate (analisi del movimento e del suono). Il modulo è quindi progettato per combinare e confrontare questo tipo informazioni.

Per ogni feature estratta, i vari stati *stable* vengono ignorati: il rilevamento di uno stato *stable* non produce alcuna modifica alle condizioni interne del sistema. La figura 4 illustra le combinazioni disponibili in relazione a ciascuno stato *high-low*. Attraverso un'interfaccia grafica utente (GUI) è possibile selezionare manualmente le combinazioni che il modulo deve rilevare durante la performance. La Figura presenta una possibile selezione delle combinazioni



**Figure 3.** Lo scheletro di un musicista estratto tramite l'unità per l'analisi del movimento. La quantità di informazione viene ridotta fino ad ottenere l'area del triangolo testa/gomiti.

di stati spesso utilizzate dall'autore durante l'esecuzione di InMuSIC.

Una volta scelta una combinazione specifica (ad esempio low energia cinetica e low loudness), il modulo verifica costantemente se i due stati vengono rilevati. Contemporaneamente e a ciascuna combinazione selezionata viene applicata una semplice condizione booleana. In breve, durante la performance, i dati inviati al modulo decisionale definiscono (i) quale condizione selezionata è attualmente vera e (ii) il tempo associato alla persistenza di ciascuna condizione verificata.

Il calcolo dei vari stati high-low consente la raccolta di informazioni relative alla variazione nel tempo delle caratteristiche estratte permettendo di definire se una feature specifica sta aumentando o diminuendo. La combinazione e il confronto degli stati high-low associati alle varie caratteristiche è quindi concepito come un ulteriore livello di astrazione all'interno dell'analisi espressiva del performer.

### 3.4 Strategie decisionali

Questa unità rappresenta il nucleo del sistema, essa infatti gestisce il livello compositivo più astratto responsabile dei processi che permettono l'interazione del sistema con il performer. La principale funzione di questo modulo è relativa alla generazione ed elaborazione di nuove informazioni musicali che possano essere da stimolo per la continuazione dell'improvvisazione.

Il modulo è realizzato per poter attuare processi decisionali associati a due differenti finestre temporali: 2-4 secondi (re-azioni in tempo reale) e 2-5 minuti (re-azioni in tempo formale). Il paradigma di riferimento si basa su ricerche relative allo studio del funzionamento della memoria uditiva degli esseri umani (memoria a breve e lungo termine) [16]. Da questo punto di vista il tentativo è quindi quello di trasferire nel sistema una sorta di consapevolezza della dimensione temporale. In concreto, sulla base del settaggio effettuato prima che la performance abbia inizio, il modulo effettua una costante correlazione tra le informazioni interpretate dalle unità di analisi audio e di movimento. Le condizioni che occorrono più frequentemente (e.g. simultanea rilevazione di alta presenza di energia cinetica e alta densità di eventi sonori) vengono, durante l'improvvisazione, memorizzate.

Questo database di carattere adattivo è la principale sorgente per lo sviluppo di processi decisionali connessi alla

finestra temporale più ampia. Allo stesso tempo, l'unità osserva costantemente i singoli parametri analizzati: questa procedura permette l'attuazione di un secondo processo parallelo che influenza la generazione ed elaborazione dei dati di controllo associati alle interazioni sonore relative alla dimensione del tempo reale.

In un certo qual modo, le varie analisi audio e di movimento vengono in questa unità software (Max/MSP) interpretate una seconda volta al fine di generare ed organizzare nel tempo i nuovi materiali sonori. Imitazione, variazione, contrasto, continuazione, interruzione, autonomia ed aleatorietà sono alcune delle metafore sulla quale si basano i processi decisionali implementati. In definitiva la procedura fondamentale riguarda la possibilità di generare nuovi elementi musicali osservando i dati in ingresso da diversi punti di vista: distinte finestre temporali e differenti tratti espressivi.

### 3.5 Generazione del suono

L'unità di generazione audio (piattaforma Max/MSP) è concepita per produrre materiali audio eterogenei. Infatti, in relazione a diversi contesti musicali, la varietà timbrica e di articolazione sonora risulta essere un requisito importante per lo sviluppo di interventi musicali stimolanti ed efficaci. I processi di elaborazione digitale del suono (Digital Sound Processing - DSP) relativi alla generazione delle proposte/risposte del sistema possono essere suddivisi in tre categorie: sintesi (generazione di suoni elettronici ex novo), sampling (elaborazione in tempo reale di suoni pre-registrati) e live processing (elaborazione in tempo reale dei materiali sonori provenienti dal microfono).

Ad ogni categoria sono associati molteplici algoritmi i quali, durante l'improvvisazione, possono essere eseguiti simultaneamente. Le singole tecniche utilizzate per la generazione ed elaborazione del suono possono essere rappresentate come voci del sistema. Ogni voce si distingue per specifiche qualità spettrali/morfologiche (distribuzione e variazione della energia nello spettro) e gestuali (articolazione e trasformazione dei materiali sonori nel tempo). Parallelamente ogni singolo algoritmo è stato progettato per poter garantire un certo grado di apertura/indeterminatezza in relazione alla generazione dei relativi materiali audio. Al fine di chiarire questo concetto, può essere utile tracciare una analogia con lo sviluppo delle voci (soprano, contralto, tenore e basso) tipico della forma del corale. Nella prassi, tenendo in considerazione i vari registri ed i moti delle singole linee, viene consentito ad ogni voce di variare in relazione a specifici gradi di libertà. Allo stesso modo vengono qui definiti dei limiti entro i quali i parametri dei vari algoritmi possono operare. Generalmente, il controllo dei singoli parametri, all'interno dei confini prestabiliti, viene di volta in volta assegnato a differenti sorgenti (e.g. dati estratti in tempo reale dalle analisi di suono e movimento, processi stocastici, informazioni memorizzate durante la performance).

L'obiettivo è quello di definire algoritmi di generazione audio che, all'interno di contesti sonori predefiniti (e.g. ambiti spettrali, dinamici e temporali) possano sviluppare continue variazioni e manipolazioni dei materiali. In altre

|                   |      | Loudness |     | Events Density |     | Pitch Deviation |     |
|-------------------|------|----------|-----|----------------|-----|-----------------|-----|
|                   |      | high     | low | high           | low | high            | low |
| QOM               | low  |          | ✓   |                |     |                 |     |
|                   | high |          |     | ✓              |     |                 |     |
| Contraction Index | low  |          |     | ✓              |     |                 |     |
|                   | high | ✓        |     |                |     |                 | ✓   |
| Smoothness        | low  |          |     |                |     |                 |     |
|                   | high |          | ✓   |                |     |                 | ✓   |

**Figure 4.** Le possibili combinazioni relative alle analisi di suono e movimento. Le combinazioni segnate sono quelle usate spesso dall'autore durante le sessioni di improvvisazione con il sistema.

parole, ogni singola voce viene concepita come in grado di esplorare spazi sonori pre-composti e il musicista è invitato a navigare questi spazi timbrici [17] in stretta collaborazione con il sistema. Questo approccio, influenzato dalle procedure tipiche della composizione algoritmica, permette, in parte, di risolvere una dicotomia che emerge nel tentativo di unire le pratiche di composizione (definizione ed organizzazione di strutture musicali - azione svolta in un tempo compositivo) ed improvvisazione (processo che porta allo sviluppo di forme musicali tramite interazioni sonore - azione svolta nel tempo performativo) nelle loro accezioni tradizionali.

#### 4. CONCLUSIONI

InMuSIC è un sistema interattivo multimodale concepito per essere utilizzato in un contesto inerente l'improvvisazione elettroacustica. L'improvvisazione musicale viene considerata come un atto espressivo spontaneo il quale implica abilità cognitive e tecniche veicolate da atteggiamenti motori e comportamenti sonori. Attraverso la correlazione delle analisi relative ai movimenti e ai suoni prodotti dal musicista, il software genera ed organizza in tempo reale materiali sonori eterogenei. InMuSIC è accordato per essere sensibile ad uno specifico vocabolario di sonorità e movimenti relativi alla pratica strumentale del clarinetto. Futuri sviluppi del sistema potrebbero prevedere la possibilità di ampliare questo vocabolario al fine di sperimentare l'utilizzo della piattaforma con differenti strumenti musicali.

Un ulteriore proposito, di carattere ben più arduo e complesso, sarebbe quello di lavorare con piccoli ensemble. In questo caso il livello di complessità concettuale ed implementativa del sistema aumenterebbe esponenzialmente poiché, oltre ad osservare il comportamento del singolo, la piattaforma dovrebbe agire tenendo in considerazione le dinamiche del gruppo. Sarebbe quindi necessario inserire nel modello compositivo presentato un ulteriore livello concettuale al fine di gestire processi decisionali in dialogo con l'atteggiamento espressivo e la condotta musicale dell'ensemble. Questo scenario, sebbene presenti sfide tecniche e compositive considerevoli, rappresenta un percorso di ricerca particolarmente stimolante.

InMuSIC è un sistema costruito intorno ad un complesso insieme di interazioni musicali che, influenzate dal comportamento del musicista, danno luogo ad un dialogo sonoro. Durante la sessione d'improvvisazione, il sistema

bilancia diversi gradi di adattività, autonomia e imitazione/variazione in relazione alle informazioni analizzate. La modalità performativa qui concepita presume quindi lo sviluppo di un intenso scambio musicale tra il sistema e il musicista. Da un punto di vista artistico, l'obiettivo/sfida dell'opera consiste nel riuscire a concretizzare una illusoria percezione di intima e stretta collaborazione tra i due agenti.

InMuSIC potrebbe essere definito come un sistema/macchina che, consentendo interventi esterni, compone/improvvisa musica attraverso una modalità dialogica. In realtà, l'oggetto di questa ricerca è in profonda relazione con quelle dinamiche umane che, all'interno di un determinato contesto culturale, influenzano sia la percezione del mondo esterno che l'espressione della propria identità. Questa ricerca pone domande relative alle modalità di generazione e trasmissione di potenziali contenuti musicali all'interno di un processo creativo spontaneo e condiviso. L'interesse quindi non è tanto indirizzato a comprendere se un computer può essere musicalmente intelligente, quanto ad effettuare una circoscritta esplorazione dell'espressività umana nel contesto musicale dell'improvvisazione elettroacustica.

#### 5. REFERENCES

- [1] G. E. Lewis, "Too many notes: Computers, complexity and culture in voyager," *Leonardo Music Journal*, vol. 10, pp. 33–39, 2000.
- [2] F. Pachet, "The continuator: Musical interaction with style," *Journal of New Music Research*, vol. 32, no. 3, pp. 333–341, 2003.
- [3] D. Van Nort, J. Braasch, and P. Oliveros, "A system for musical improvisation combining sonic gesture recognition and genetic algorithms," in *Proceedings of the 6th Sound and Music Computing Conference*, pp. 131–136, 2009.
- [4] S. Jordà, "Instruments and players: Some thoughts on digital lutherie," *Journal of New Music Research*, vol. 33, no. 3, pp. 321–341, 2004.
- [5] M. Leman and A. Camurri, "Understanding musical expressiveness using interactive multimedia platforms," *Musicae Scientiae*, vol. 10, no. 1 suppl, pp. 209–233, 2006.
- [6] J. Chadabe, "The limitations of mapping as a structural descriptive in electronic instruments," in *Proceedings of the 2002 conference on New interfaces for musical expression*, pp. 1–5, National University of Singapore, 2002.
- [7] M. Malt and E. Jourdan, "Real-time uses of low level sound descriptors as event detection functions using the max/msp zsa. descriptors library," *Proceedings of the 12th Brazilian Symposium on Computer Music*, 2009.

- [8] A. De Cheveigné and H. Kawahara, “Yin, a fundamental frequency estimator for speech and music,” *The Journal of the Acoustical Society of America*, vol. 111, no. 4, pp. 1917–1930, 2002.
- [9] T. Jehan and B. Schoner, “An audio-driven perceptually meaningful timbre synthesizer,” *Analysis*, vol. 2, no. 3, p. 4, 2002.
- [10] A. Camurri, G. Volpe, G. D. Poli, and M. Leman, “Communicating expressiveness and affect in multimodal interactive systems,” *Ieee Multimedia*, vol. 12, no. 1, pp. 43–53, 2005.
- [11] A. Camurri, S. Hashimoto, M. Ricchetti, A. Ricci, K. Suzuki, R. Trocca, and G. Volpe, “Eyesweb: Toward gesture and affect recognition in interactive dance and music systems,” *Computer Music Journal*, vol. 24, no. 1, pp. 57–69, 2000.
- [12] G. Kurtenbach and E. A. Hulteen, “Gestures in human-computer communication,” *The art of human-computer interface design*, pp. 309–317, 1990.
- [13] E. Douglas-Cowie, N. Campbell, R. Cowie, and P. Roach, “Emotional speech: Towards a new generation of databases,” *Speech communication*, vol. 40, no. 1, pp. 33–60, 2003.
- [14] A. Camurri, I. Lagerlöf, and G. Volpe, “Recognizing emotion from dance movement: comparison of spectator recognition and automated techniques,” *International journal of human-computer studies*, vol. 59, no. 1, pp. 213–225, 2003.
- [15] D. Glowinski, N. Dael, A. Camurri, G. Volpe, M. Morcillaro, and K. Scherer, “Toward a minimal representation of affective gestures,” *Affective Computing, IEEE Transactions on*, vol. 2, no. 2, pp. 106–118, 2011.
- [16] B. Snyder, *Music and memory: An introduction*. MIT press, 2000.
- [17] D. L. Wessel, “Timbre space as a musical control structure,” *Computer music journal*, pp. 45–52, 1979.



