UNIVERSITY OF UDINE

PhD Course in Multimedia Communication

XXV Cycle

Interactive and Multimodal Environments for Learning

A thesis submitted for the degree of

Philosophiæ Doctor

PhD Student: Serena Zanolla

Supervisor: Sergio Canazza Targon

April 2013

Serena Zanolla

Department of Mathematics and Computer Science, University of Udine, Italy

PhD in Multimedia Communication, Department of Human Sciences, 2010-2012, XXV Cycle

email: serena.zanolla@uniud.it

Supervisor:

Sergio Canazza Targon

Assistant Professor, Department of Information Engineering, University of Padova, Italy

Reading Committee:

Evangelos Himonides

DipMus, BSc 1st HONS, MA, PhD, MBCS CITP

Senior Lecturer in Music Technology Education, iMerc Co-ordinator

Department of Culture, Communication and Media

Institute of Education, University of London, UK

Maria Chiara Levorato

CLASTA President

Department of Developmental Psychology and Socialization

Lab-Las – Language Acquisition Studies Laboratory

University of Padova, Italy

Day of the defense: April 12, 2013

PhD Student's signature:

PhD Supervisor's signature:

Abstract

The doctoral research project, presented in this thesis, involves the design, development, and experimentation of Interactive and Multimodal Environments (IMEs) for real educational contexts. The IMEs are technological systems able to provide students with a multimodal and interactive learning experience, through sensors for data acquisition, real-time processing (hardware and software) components for input data processing, and output components for reproduction of multimedia content.

The IMEs involving physical spaces have been identified as the best solution to implement a high level of interactivity and multimodality in learning processes. The IMEs embedded in physical spaces are indeed able to guarantee full-body interaction and a multisensory experience which allow users to utilize their whole body, gestures, and their voice to fulfill a task.

These systems enable teachers to reach a heterogeneous school population arising from a significant diversification in levels of learning, a high proportion of foreign children, and the growing number of children with disabilities who, today, attend regular schools – spending at least 80% of their day in regular classrooms – and receive early intervention services. The growing number of pupils with disabilities at school is explained also by improved diagnoses that facilitate the identification of the population generally recognized as "disabled" and the population who has "special needs" related to milder forms of impairments and learning disorders.

The first IME documented in this thesis is the Stanza Logo-Motoria that this author conceived in 2008 with the aim of rethinking the educational dynamics by means of innovative teaching approaches based on technological devices. The first application of the Stanza Logo-Motoria, the Resonant Memory, was developed in collaboration with the InfoMus Lab of the Communication, Computer, and System Science Department (DIST) of Genova University. The Resonant Memory application, used at school since 2009 with scientific research purposes, was developed by the *Centro di Sonologia Computazionale* (CSC) of the Information Engineering Department (DEI) of Padova University.

Moreover, this thesis documents Fiaba Magica and PlayAndLearn, two applications for the Stanza Logo-Motoria, that this author conceived during the PhD period with the aim of supporting, respectively, a strengthening path of gestural intentionality of children with multiple-disabilities and teaching new vocabulary and meanings. The Fiaba Magica software was developed by the University of Udine and then, improved by CSC of Padova University; the PlayAndLearn software was developed by CSC of Padova University.

Further IMEs, focusing on specific educational objectives, were conceived by this author during the PhD period: (a) Memory Sonoro, a system for people with severe visual impairment aimed at developing tactile-auditory perception integration, and (b) SoundRise, aimed at teaching the acoustic properties of the voice. The Memory Sonoro and SoundRise applications were developed, respectively, by the Mathematics and Computer Science Department (DIMI) of Udine University and the CSC of Padova University.

Finally, this doctoral thesis presents two more systems, SoundingARM (Acoustic Representation of a Map) and ParrotGameDiscrimination, that this author contributed to their implementation. The SoundingARM system was designed and developed in collaboration with the CSC of the Information Engineering Department of Padova University within the project "*Evoluzione dei sistemi di feedback acustico per la prima domiciliazione dei soggetti con inabilità visiva acquisita*" (Evolution of acoustic feedback systems for the homecoming of subjects with acquired vision impairment) funded by the Friuli Venezia Giulia Region[1] (Italy). ParrotGameDiscrimination, of which this author is conducting an experimentation in several educational institutions, is a multimodal video-game for mobile devices that was developed by the CSC of Padova University in collaboration with the Department of Developmental Psychology and Socialization (DDPS) of the Padova University in order to enhance the assessment techniques of pupil's auditory skills.

In order to measure the educational effectiveness of these IMEs, several usability tests and evaluations were carried out. The results of the usability tests highlighted the weak points, allowing solutions to be found and the performance of the systems to be improved.

This thesis is organized as follows; the Introduction (Chapter 1):

1. outlines the evident need for a rethinking of educational dynamics by implementing innovative teaching methods based on technological devices;

[1] Decree n. 1265/AREF dd. November 25th, 2010

2. briefly introduces the Stanza Logo-Motoria and its educational applications (Resonant Memory, Fiaba Magica, and PlayAndLearn);
3. presents the IMEs (SoundingARM, Memory Sonoro, SoundRise, and ParrotGameDiscrimination) developed during the PhD period.

The theoretical framework of the educational employment of IMEs in physical spaces and related works are addressed in Chapter 2.
Chapter 3 fully documents the Stanza Logo-Motoria system and its specific applications. The preliminary study of the Stanza Logo-Motoria in Resonant Memory modality is presented in Chapter 4.
Chapter 5 explains SoundingARM (5.1) and Memory Sonoro (5.3), IMEs specifically tailored for people with visual impairment, and addresses their theoretical foundations and related works.
Chapter 6 outlines the theoretical framework, related works, and detailed description of SoundRise (6.2) and ParrotGameDiscrimination (6.3), IMEs for speech production/discrimination.
Finally, Chapter 7 presents a general discussion on the research, the conclusions, and future potential developments for the ongoing project.

To me.

Acknowledgements

First, I would like to express my gratitude to Veronica Zoratto, who was my inspiration for the idea of the Stanza Logo-Motoria.

My gratitude to my supervisor, Sergio Canazza Targon, and Antonio Rodà, Assistant Professors at the Engineering Information Department (DEI) of Padova University (Italy) who have supported my ideas and my work, since the very beginning of my research.

I would like to express my gratitude to InfoMus Lab of Genova University (Italy), in particular, Antonio Camurri, Corrado Canepa, Paolo Coletta, and Gualtiero Volpe; I am grateful for their precious collaboration during the preliminary development of the Stanza Logo-Motoria.

My gratitude to Silvia Steppi Zanin, headmistress of the "Gorizia 1" Comprehensive Institute (Italy), who made the installation of the first Stanza Logo-Motoria at the "E. Frinta" Primary School possible. I owe my deepest gratitude to Fabrizia Perco, assistant head of the "Gorizia 1" Comprehensive Institute, for the trust, impeccable support, and patient collaboration in my work. I would like to express my gratitude to my colleagues of the "Gorizia 1" Comprehensive Institute who were always kind, helpful, and patient with me during all the experimental period. I would like to offer my gratitude and my admiration to all the school children who participated in the three-year experimentation of the Interactive Multimodal Environments (IMEs); school children are the main source of inspiration in my daily work.

I want to thank Dario Faganel for his fundamental contribution in preparing the classroom for the installation of the Stanza Logo-Motoria at the "E. Frinta" Primary School of Gorizia (Italy); I am grateful to Maurizio Grusovin – in memoriam – who installed the electrical system of the room.

I would not be able to support all the challenges of these years without all my friends, in particular, I have no words to describe my gratitude to Enza Scalone, who provided me strength, calm, and sense of focus during critical phases of my academic and personal life and to Maura Tomasi for patience and competence in proofreading my papers and the PhD thesis. My gratitude to Pier Paolo Lenaz, vice-president of the Regional Institute

"Rittmeyer" for Blind of Trieste (Italy) for his support in specific insights of my research.
I thank my reading committee, Evangelos Himonides (Senior Lecturer at the Institute of Education, University of London, UK) and Maria Chiara Levorato (CLASTA President, Department of Developmental Psychology and Socialization, University of Padova, Italy) for their valuable advice and detailed comments.
Finally, I express my deepest gratitude to my family.

Contents

1

Introduction

1.1 Interactive and Multimodal Environments (IMEs) for Learning

The aim of this doctoral research project involves the design, development, and validation of Interactive and Multimodal Environments (IMEs) for learning. The IMEs are hardware and software systems able to provide users with multimedia feedback, according to their movements, gestures, and sounds (i.e. voice). The IMEs require an acquisition system integrated with processing components able to interpret the measurements of the sensors, all in real time. Usually, an IME has the following main components:

- *Input Components*, sensors (e.g. a webcam, a microphone) for tracking the full-body movements, gestures, speech, and sounds;

- *Real-Time Processing Components*, a control unit (hardware and software) for the processing and mapping of the data output deriving from Input Components (i.e. users' actions/sounds);

- *Multimedia Output Components*, such as loudspeakers, headphones, and projectors to reproduce audio/visual content.

1. Introduction

The focus of the research is on IMEs which convey multimedia contents in the physical space. This kind of IMEs is embedded within a physical environment which can be a classroom, a museum, a stage, etc. Involving a physical environment increases the possibilities of interaction with the real world; this can be augmented by verbal and non-verbal contents. The IME in a physical space allows users to interact with multimedia content in real time.

The scope of designing and developing IMEs for learning is based on the assumption that any educational context would probably need to be renewed upon the introduction of new educational tools and methodologies that are able to offer a high-level of interactivity and multimodality.

The need for a methodological rethinking is due to the fact that, nowadays, teaching occurs in a highly heterogeneous educational context: there is a significant diversification in levels of learning, a high proportion of foreign children, and a growing number of children with disabilities who attend regular schools (96%), spending at least 80% of their day in regular classrooms. Moreover, not only 2% of the population – generally recognized as "disabled" – is identified but also 8% who have "special needs" – related to milder forms of impairments and learning disorders – thanks to improved diagnosis ([1], [2] and [3]). This heterogeneity would entail a major transformation in the way our schools are organized. It suggests that teachers need to be trained to present their lessons in a wide variety of ways using music, cooperative learning, art activities, role play, multimedia, and technologies, in order to recognize and nurture all the cognitive styles, to encourage interaction with the world, the general development of the person, and the achievement of the highest possible level of learning ([4], [5], and [6]).

Today, Information and Communication Technologies (ICTs) are recognized as useful instructional tools [7]. ICTs indeed support teachers by offering a variety of technological solutions, such as computers, interactive whiteboards, and wireless tablets which can supply pliability and creativity in the processes of learning/teaching. Nevertheless, these ICTs lack the involvement of a physical environment, which is essential in education, since a critical part of a children's cognitive development is believed to be based upon their interaction with the physical world, within a social and cultural environment [6].

The main idea is to bring "physicality" into "brains-on" learning, combining full-body movements and gestures within a broad physical space, with higher cognitive activities like thinking, reasoning, and reflecting [8]. This approach is based on a fundamental assertion offered by contemporary developmental theories: effective learning occurs when meaning is obtained from experience with the things of the world [9; 6]. The IMEs, able to provide such interaction, enable the user to exploit the whole body postural space to fulfill a task. Consequently, learning becomes more engaging since knowledge is obtained through a motor experience.

1.1.1 The Stanza Logo-Motoria

The Stanza Logo-Motoria was conceived by this author in compliance with the aims described above [10]. This system, fully documented in Chapter 3, has an originating educational objective: to have a classroom transform into an IME in order to provide schoolchildren with alternative learning experiences, therefore enabling them to respond to their real educational needs. On the other hand, the Stanza Logo-Motoria system aims at implementing new dynamics of teaching in order to help teachers face the heterogeneity of the educational context. Offering a multimodal access to knowledge and a high-level of interaction with others and the environment, the Stanza Logo-Motoria enables the implementation of alternative/additional teaching methodologies.

The Stanza Logo-Motoria is a modular system with which a real environment is "augmented" with sounds and images by means of video-based tracking techniques. The Stanza Logo-Motoria is a ample empty space (namely a classroom, a gym, a stage) where the user's presence and movements trigger the playback of audio/visual content.

A basic configuration of the Stanza Logo-Motoria system is implemented by:

standard sensors (e.g. a webcam positioned on the ceiling) which detect the user's full body movements and gestures;

a software for multimodal interaction which a) processes the video signal in order to extract the low-level motion features (position of the user in space and quality of gestures) and b) controls the interaction logic and the real-time rendering of audiovisual contents;

peripheral devices such as loudspeakers and projectors for the reproduction of multimedia content.

The modularity of the Stanza Logo-Motoria system allows the implementation of diversified levels of interactivity and multimodality according to specific educational needs. The system indeed allows the operator to use different sensors, various types of peripheral devices, or diverse software applications. Therefore, in order to provide the system with different interaction logics, several educational applications have been developed: Resonant Memory, Fiaba Magica, and PlayAndLearn which are briefly described below (Sect. 1.1.1.1).

Moreover, in order to fulfill more specific educational needs of people with severe visual impairments and pupils with language disorders, other interactive and multimodal systems have been developed: SoundingARM, Memory Sonoro, SoundRise, and ParrotGameDiscrimination. These systems are concisely presented in paragraph 1.2.

1.1.1.1 Educational Applications of the Stanza Logo-Motoria

Since 2009, the Stanza Logo-Motoria is permanently installed at the "E. Frinta" Primary School ("Gorizia 1" Comprehensive Institute, Gorizia, Italy). To date, several applications of the Stanza Logo-Motoria have been developed; the first is Resonant Memory with which the space, acquired by a webcam, is subdivided in nine areas: eight peripheral and one central. The user's presence within a specific area triggers the audio reproduction of the corresponding audio content. Noises, music and environmental sounds are synchronized to the peripheral areas, whereas the central area is synchronized with an audio reproduction of the contents to be taught (e.g. a story) that contain the elements to be connected with the sounds positioned in the various peripheral areas. The child, who is free to move in space and does not need to be wired onto any type of sensor, carefully listens to the content reproduced in the central area, reaches the different peripheral areas experimenting with the sounds and finally, in a fun and effective manner, "composes the soundtrack" of the lesson [10; 11; 12].
Resonant Memory was conceived by this author in 2008 – under the supervision of Sergio Canazza Targon – in collaboration with the InfoMus Lab of the Communication, Computer, and System Science Department (DIST) of Genova University (Italy) who developed the first patch based on the EyesWeb XMI platform[1]. In 2009, Antonio Rodà, currently Assistant Professor at the Information Engineering Department (DEI) of Padova University, developed the second version of the Resonant Memory patch suitable for the school environment and useful for scientific research purposes.
The Resonant Memory application is used to create a great deal of educational activities involving several school subjects:

1. as a tool to activate the written production of a tale;
2. in order to study a school subject such as History;
3. as a tool to develop the spatial ability on the part of students with severe visual impairment;
4. as a method to improve the communication skills of foreign children who are learning Italian as their Second Language;
5. as a listening tool for ESL (English as a Second Language) classes.

On the basis of the last activity listed above (ESL classes), a preliminary study of the Resonant Memory application was conceived and carried out, following a validation protocol (Chapter 4). For two school years (2011 and 2012), from February to June, a pre- and post-test non-equivalent group design [13; 14]

[1] http://www.infomus.org

was applied. The validation protocol implied the comparison of the pre-experimentation situation with the final performance (listening and comprehension skills) of two comparable classes, in order to verify the experimental hypothesis: whether pupils using the Stanza Logo-Motoria, in Resonant Memory modality, make more progress in ESL listening comprehension of words and sentences, thanks to the experimental factor (the innovation introduced), than those who perform listening activities in the language laboratory. With this aim in mind, in each experimental phase, a pre and post-listening test were administered to both classes. The interesting results of this preliminary study are fully presented in Chapter 4 and published in [12].

A further application designed for the Stanza Logo-Motoria is Fiaba Magica (Fairy Tale) which can be used to improve the residual movement and the gesture intentionality of pupils with severe impairments. This application enables the user to reconstruct sounds and images of a tale by performing a simple gesture such as raising their arms sideward and moving within a specific space.

In particular, when using Fiaba Magica a child enters a specific area of the Stanza Logo-Motoria and activates a) the audio reproduction of the first part of the story and b) the projection on the screen of the corresponding image (for example two characters). Once the audio reproduction of the first part of the story is played, by lifting their arms sideways, the user can play with the characters projected, animating them. Moving to the next area, the child activates a) the projection of a new sequence of the story which includes another set of characters and b) the audio reproduction of the narrative sequence itself. The third advancement in the story is performed in the same way as the other two [12].

In collaboration with parents and teachers, in February 2011, a two-day preliminary usability testing of the Fiaba Magica application was carried out; the performances of two pupils with severe impairments were assessed. This preliminary evaluation of the system highlighted the fact that both pupils were motivated to walk, to lift their arms sideways, to hold their head up, to control their balance in standing and walking, and to pay attention [11].

The PlayAndLearn application is an evolution of Fiaba Magica. It is an interactive and multimodal application with which pupils may acquire new vocabulary and meanings. The user's movements and gestures trigger audio/visual contents. As in Fiaba Magica, here too the space is subdivided into three areas; the presence of the user within each area is synchronized with a) the audio reproduction of a spoken word and its meaning and b) the projection of two animations/images. Firstly, the user listens to the audio reproduction of a word and then selects the animation which corresponds to the word heard. The selection is performed by widening their right or left arm sideward according to the position of the animation on the screen. The PlayAndLearn application tracks the user's movements and gestures by means of the Microsoft Kinect sensor.

At the time of writing, a validation protocol of the PlayAndLearn application is being planned: a between-subjects experimental design will be applied to a First class at primary school level. This preliminary study will aim at verifying whether vocabulary acquisition obtained by means of the PlayAndLearn application is more effective than vocabulary acquisition obtained by a traditional PowerPoint presentation.

1.2 Interactive and Multimodal Environments for Specific Objectives

Another important aim of this doctoral research project is the design, development, and experimentation of IMEs as assistive aids to meet specific educational needs. Following this intention, SoundingARM [15] and Memory Sonoro [16], IMEs specifically tailored for people with visual impairments (detailed in Chapter 5), SoundRise and ParrotGameDiscrimination, IMEs for speech production/discrimination (detailed in Chapter 6), have been designed, developed, and experimented with.

The overall goal of these IMEs is to provide users with a visual and/or acoustic feedback that derives from the analysis of their gestures and speech. These IMEs are indeed able to generate multimedia feedback that supports sensory compensation, the evaluation process on the part of the operator (teacher, therapist), and the acquisition of specific skills. Within these environments, the users' actions and sounds (vocal output) are the point of focus, enabling the participant to actively engage with the content provided by the system.

1.2.1 IMEs for Visual Impairments

The SoundingARM system [15] aims at offering blind users with an auditory "first sight" of space, an acoustic map of space, allowing them to immediately be aware of the type of environment and the objects it contains. The SoundingARM system provides blind people with a means to explore a room by standing at the threshold and performing a simple gesture: pointing. When the user points their arm around to find an object, the system promptly answers supplying an audio feedback. Moreover, the user obtains the information regarding the position of an object through the gesture itself: the object is located in the direction in which the finger points.

SoundingARM does not require wearing sensors, handling pointing devices, or any tag to mark the environment. On the contrary, it uses standard hardware such as the Microsoft Kinect sensor, a normal Windows 7 computer, and a software to manage the localization data in order to generate vocal information on the space.

Three usability tests were conducted in order to evaluate: a) the system performance; b) the consistency of the spatial map; and c) the use of Auditory Icons versus the use of Speech Icons. The results of the usability testing are fully reported in Chapter 5.

The SoundingARM system has been designed and developed in collaboration with the *Centro di Sonologia Computazionale* (CSC) of the Information Engineering Department of the Padova University within the project "*Evoluzione dei sistemi di feedback acustico per la prima domiciliazione dei soggetti con inabilità visiva acquisita*" (Evolution of acoustic feedback systems for the homecoming of subjects with acquired vision impairment) funded by the Friuli Venezia Giulia Region[1] (Italy).

The ability to orientate oneself and move within a space is the result of a specific sensorial education which allows people with visual impairment to fully exploit the sensorial stimuli and use them to explore the environment, moving safely, understanding the position, and the spatial proportion of the objects. The Memory Sonoro system [16] finds application exactly in this type of context, positioning itself as a teaching aid for the blind, for the development and exploitation of the vicarious senses and the reduction of the secondary effects of blindness. Memory Sonoro was conceived by this author during the PhD period and developed in collaboration with the Mathematics and Computer Science Department (DIMI) of Udine University: it is an interactive multimodal system that can be used by the blind and partially sighted, as it allows them to play a tactile-audio memory game "Memory". With Memory Sonoro, the classical interaction modality, mainly based on sight, has been replaced by a) hearing (with the reproduction of spatialized sounds) and b) touch (making use of tactile interfaces).

The system has been tested on a heterogeneous sample of 12 users by means of a preliminary usability test. The results, fully reported in Chapter 5, have evidenced different modalities of use from those imagined during the engineering phase and a couple of limitations both in terms of architecture and implementation of the system.

1.2.2 IMEs for Speech Production/Discrimination

The use of voice as an enactive instrument for learning is particularly interesting: the voice is the universal instrument for transmitting information (both verbal and non-verbal) since it includes paralinguistic elements that convey emotions and moods ([17] and [18]). Indeed, the speech perception involves the ability to identify the spectral content and the amplitude and fundamental modulation. This is particularly important when the user, because of physical or cognitive impairments, is able to communicate exclusively through simple vocal modulations. This assumption triggered the design and the development of SoundRise.

[1] Decree n. 1265/AREF dd. November 25th, 2010

SoundRise was conceived by this author during the PhD period and developed by the CSC of the University of Padova; it is an interactive and multimodal application based on speech recognition technologies, able to provide graphical feedback according to each feature of the vocal signal. The goal of SoundRise is to teach pupils the vocal parameters and how to control them. This application can be used by means of both a computer or the Stanza Logo-Motoria.

The SoundRise application creates an amusing graphical landscape on a screen; a yellow sun appears above the horizon when users make a vocal sound. By selecting specific buttons (for the computer version) or a specific area of the Stanza Logo-Motoria (in the Stanza Logo-Motoria version), users can choose which voice feature to visualize. For instance, if the user selects the amplitude parameter, the system analyzes the amplitude of sound and produces a variation in the size of the sun according to the change in sound pressure levels. The voice-based activities provided by SoundRise allow to raise learners' awareness in terms of potentialities of their voice and its attributes.

The SoundRise application has been experimented with children and adults by means of two usability tests in order to evaluate a) the system performance and b) whether the interface is intuitive for the user. The first usability test has allowed to detect that the system did not fulfill the task of the vowel identification function and to identify the reason. The second test allowed to observe that timbre is associated with the less intuitive graphical feature and the reason why. The results of this usability testing are fully reported in Chapter 6.

The ability to correctly interpret utterances produced by another person is essential for interpreting social interactions. Besides the difficulties with the possible social interactions, children with auditory processing related conditions are likely to experience severe difficulties in the class environment. On the basis of this assumption, ParrotGameDiscrimination (Chapter 6), a multimodal video-game for mobile (tablet, PC, in particular), has been developed by the CSC of the University of Padova in collaboration with the Developmental and Socialization Psychology Department (DPSS) of the University of Padova. ParrotGameDiscrimination has been designed in order to measure, for each sound feature, the child's ability to discriminate the differences among a number of sound stimuli (sinusoidal as well as complex sounds), synthesized with different parameters (timbre, pitch, intensity level). The main goal of this application is to widen the scientific research about the auditory processing of speech in specific populations of pupils (such as children with Down's Syndrome) who, due to auditory processing disorders, encounter difficulties in speech production. Moreover, this system allows the study, assessment, and intervention in pathologies and disorders of various etiology that determine an impairment in speech production and articulation.

At the time of writing, this author is conducting an experimentation of ParrotGameDiscrimination in several educational institutions in order to evaluate the speech discrimination skills of preschoolers. Figure 1.1 shows the IMEs developed during the PhD period. Each IME illustrates: the type of input data (i.e. visual or language); type of user's interaction; type of output data (i.e. auditory and/or visual); the aims (learning, rehabilitation, or assessment); and the use (application and/or research).

		Input data				User's interaction					Output data		Aim			Use	
		Video	Vocal	Touch Screen	Infrared	Visual	Tactile	Vocal	Motion	Space	Audio	Video	Learning	Rehabilitation	Assessment	Application	Research
SLM	RM	X							X	X	X		X	X		X	X
	FM	X							X	X	X	X	X	X		X	X
	P&L	X				X			X	X	X	X	X	X		X	X
MS		X					X				X		X			X	X
SARM					X				X	X	X		X	X		X	X
SR			X			X		X				X	X			X	X
PGD				X		X	X				X	X			X	X	X

Legend
SLM: Stanza Logo-Motoria
RM: Resonant Memory
FM: Fiaba Magica
P&L: PlayAndLearn
MS: Memory Sonoro
SARM: SoundingARM
SR: SoundRise
PGD: ParrotGameDiscrimination

Figure 1.1: IMEs - The interactive and multimodal systems developed during the PhD period.

2

Interactive and Multimodal Learning Environments in Physical Spaces

2.1 Theoretical Framework

As briefly mentioned in the Introduction (Chapter 1), this thesis is concerned with the design, development, and validation of Interactive Multimodal Environments (IMEs) for learning.

IMEs are technological systems capable of establishing multimodal user interaction by providing real-time feedback in terms of sound, music, and visual media. From the Artificial Intelligence (AI) perspective, an IME consists of hardware and software components that supply multimedia feedback according to the recognition of a user's movements, gestures, and speech.

Generally, an IME is constituted by: a) *Input Components* (sensors) for tracking the full-body movements, gestures, speech, and sounds; b) *Real-Time Processing Components*, a control unit (hardware and software) for processing and mapping of the user's actions/sounds; c) and *Output Components* (loudspeakers, projectors,..) for reproducing audio/visual contents. Moreover, the IME needs to have certain specific features [19]:

- *real-time, in-time, perceptual time*; be able to react within the documented time thresholds of the human perceptual system (0.1 second [20] in order to provide the sense of "real time"; the user perceives system reactions as immediate);

- *interactivity*, provided by multimodal interfaces which involve gestures, voice, sound, music, and vision;
- *a non-invasive sensor system* which allows the user to focus on creative activities and not on technology.

The main aim of an IME is to extend physical reality, without deceiving the sensorimotor human system, to allow the user interact with the real world "augmented" by technology. Unlike a musical instrument, the IME is pliable and modifiable by means of the user's behavior [19]. An interesting incentive to use IMEs is to adapt new technologies to basic human forms of communication including those which are nonlinguistic and based on movement: actions, gestures, and mimetic activities [21].
By means of technology, it is now possible to "augment" the traditional learning environment (e.g. a classroom) with a great amount of interactivity which could be utilized to better deal with the real needs of school children [22]. The school staff, indeed, often face cases of learning difficulties, problematic relationships, and school failures, experiencing an overall sense of helplessness that also arises from the lack of specific criteria on how to deal with these issues provided by the traditional methodologies. Moreover, today there is the awareness that teaching must be more student-centered and learning does not automatically occur when information is presented to a child; on the contrary, active involvement, interacting, and part-taking, are central activities for learning effectiveness ([23], [24], [25]).
Technology should be used as a tool of knowledge construction and representation and not for mere passive teaching [7]. Nowadays, technology supports teachers by offering a variety of technological solutions carried out by computers, interactive whiteboards, wireless tablets, whose level of interactivity depend on how and to which extent children participate in the tasks [26]. However, the use of ICTs (Information and Communication Technologies) in the classroom tends to support passive and non collaborative forms of learning rather then full-body interaction. Therefore, at school, it is necessary to implement new forms of interaction that "exploit the *physical* and the *digital* in a diversity of ways that move beyond the desktop genre of interactions" [8].
Although computers, interactive whiteboards, wireless tablets can offer creative learning/teaching possibilities, they lack an important element: the involvement of a physical environment. This element is basic in an educational context, since a critical part of a child's early cognitive development involves interacting with the surrounding physical world, within a social and cultural environment [6]. The main idea is to bring "physicality" into "brains-on" learning combining full-body movements and gestures with higher order cognitive activities like thinking, reasoning and reflecting [8]. This approach is based

on a fundamental assertion offered by contemporary developmental theories: effective learning occurs when meaning is obtained by interacting with the things of world [9]. "Physical engagement with something creates an involvement and activeness in learning that passive listening or watching does not. This, in part, increases levels of motivation and an interest in the activity or learning context. High levels of engagement can in turn affect the cognitive interaction of the learner, in terms of their attention, inquisitiveness and reflection" [8].

Children usually perform various types of movement in order to explore an environment: crawling, rolling around, dancing, walking, gesturing, and jumping which can be used to trigger different digital events (e.g., sounds, animations) becoming an integral part of their learning experience. Among these, "there are three basic human motor movements: rolling, crawling/walking, and jumping which correspond with the way that information travels in the brain: side to side across the corpus callosum, back to front across the motor cortex and up and down from the bottom to the top of the brain. The brain uses its motor patterns as the framework for learning" [27].

This doctoral thesis is mainly focused on the design and the development of IMEs implementing full-body interaction which enables the user exploit the whole-body postural space in order to fulfill a task. This kind of IME requires an acquisition system integrated with processing components able to interpret sensor measurements in real time. With regards to the acquisition system, the main focus is on motion tracking systems. Several motion capture systems have been developed to track human movements; the motion tracking is generally performed thanks to sensors fixed on the user's body, or markers positioned in the environment. The technology mainly used for motion capture is: optical systems exploiting an infrared camera; magnetic systems sensitive to magnetic information instead of light information; exoskeletons using a light exoskeleton attached to the user's body; and video-based systems which recognize the user's movement from the analysis of the image. This last solution, even if costly – especially for real-time applications – in terms of computational resources, leaves the user free to move in space avoiding the use of sensors [28].

In agreement with the above cognitive and technological perspectives, the Stanza Logo-Motoria (fully documented in Chapter 3), a video-based system with which users explore the physical space without wearing sensors, and, during the PhD period, the following IMEs for learning have been designed and developed:

1. SoundRise (Sect. 6.2) and SoundingARM (Sect. 5.1), in collaboration with the CSC (*Centro di Sonologia Computazionale*) of the University of Padova;

2. ParrotGameDiscrimination (Sect. 6.3), in collaboration with the CSC of the University of Padova and the Developmental and Socialization Psychology Department (DPSS) of the University of Padova;

3. the Memory Sonoro system (Sect. 5.3), in collaboration with the Department of Mathematics and Computer Science (DIMI) of the Udine University.

Each of these IMEs implements different high-level interactional activities and provides multimedia content in order: a) to meet the real needs of students; b) to involve their different cognitive styles; and c) to help teachers deal with the heterogeneity of today's educational context. The design of all the IMEs above, is deeply rooted in other important cognition theories such as *Embodied Cognition*, *Enaction*, the *Theory of Multimedia Learning*, and the *Dual Coding* theory which are fully explained in the following sections.

2.1.1 Interactive Learning: Embodiment and Enactment

Interactivity, in the educational context, is a characteristic of learning environments which enables multidirectional communication, a two-way action opposed to a one-way action, between the learner and the instructional system. When IMEs are used for learning, their interactional component is of extreme importance. This consideration is based on the assumption that a meaningful knowledge is not propositional (knowing that) but procedural (knowing how). The theories of *Embodied Cognition* (EC) and *Enaction* provide a conceptual framework in which IMEs can be studied from a more comprehensive point of view.

The recognized viewpoint of *Embodied Cognition* (EC) posits that cognitive processes are based on body–world interaction; so, cognition is not centralized, abstract, and separated from the body but originates from sensorimotor processing. Therefore, body is not a mere container of the mind but rather contributes, molds, and enacts the construction of meaning [29]. This idea is definitively revolutionary in the realm of learning practices. The first notion of *embodied* knowledge emerged in the field of phenomenology [30] as an alternative to the Cartesian mind–body division. It evolved into the concept of *embodied cognition* or *embodied mind* [31] researched based evidence from the field of cognitive sciences, linguistics, robotics, biology, and other fields. The EC framework, from a neurophysiological point of view, is based on a mirror-system which enables the coupling of perception and action by means of shared neuronal regions in the human brain [32; 33].

For the EC approach [29]:

"Cognition is situated." Cognition, involving perception and action, occurs in the real world;

"Cognition is time-pressured." Cognition functions under the pressure of real-time interaction with the environment;

"Human beings off-load cognitive work into the environment." Due to limits in the attention and working memory, human beings use the environment to simplify cognitive tasks (e.g. counting with one's fingers or using paper and pencil to store intermediate information, etc.);

"The environment is part of the cognitive system." In such cases, the human being's actions are best understood by including tools with the body to form a single activity (for example, the blind people does not feel "with the hand holding the cane", but "with the cane"; in this case, the cane becomes a part of the body);

"Cognition is for action." The mind guides action;

"Off-line cognition is body-based." For instance, the mere observation of an action triggers the activation of neurons used in bodily perception and action ([32; 33]).

In the field of *Embodied Music Cognition*, Leman [34] offers the viewpoint that the human body is a "natural mediator between subjective experience and physical reality". The Embodied Music Cognition framework is based on theories of Enaction [35] and Entrainment [36], and the mirror neuron research [32; 33]. Therefore, from an educational point of view, the embodied cognition approach entails the idea that "learning by moving" is an activity that relies on the coupling of action-perception, and that the human body is a *medium* between subjective experiences and the physical environment.
The understanding that body interacts with its environment leads also to the concept of *enactment* ([9; 37; 35]). Enactment or *enactive knowledge* is acquired through the action of doing. The term *enaction* was first introduced by Jerome Bruner ([9; 37]) who describes three ways of organizing knowledge, which are also three developmental stages, and the corresponding forms of interaction with the world: enactive, iconic, and symbolic.

- Enactive knowledge is based on motor skills, so, enactive representations are "acquired by doing";
- iconic knowledge is based on recognition of visual information;
- symbolic knowledge derives from cognitive functions (such as language).

Each developmental phase of human beings is characterized by the use of a specific way of knowledge representation; hence, adults can count on all three types of representation. The enactive approach

[9; 35; 23] attests that much knowledge is acquired through action within an environment and it is constructed on motor skills: the representations are gained "by doing" and "doing" is the tool for learning [37]. This approach to cognition can help schoolteachers face the specific learning difficulties of pupils, given that teaching is considered as a co-construction in which the active participation of the learner is essential and where the body is used as a learning channel [25]. "Learning by doing" is a theoretical dimension which has to be considered in designing IMEs for learning where users have to acquire knowledge through action.
In HCI (Human Computer Interaction), interactivity is defined "as the extent to which users can participate in modifying the form and content of a mediated [by technology] environment in real time" [38]. Three main factors contribute to make a mediated environment interactive:

speed, the user' actions alter the environment in real time;

range, the number of action possibilities;

mapping, the way in which human actions are connected to multimedia contents within a mediated environment.

Since Sutherland's SketchPad in 1961 [39] and Xerox's Alto[1] in 1973, users have long been aware of technologies other than the traditional keyboard for interacting with a system. Recently, multimodal interaction – along with progress in field of HCI – has emerged as an active field of research [22]. In particular, the basic goal of the HCI research is to improve the user – computer interaction by making computers more user-friendly and receptive to the user's needs. The interaction is nowadays not limited to an action – reaction loop between the computer and the user: current research investigations focus on the tracking of the body parts for a well-directed reaction [28].

2.1.2 Multimedia/Multimodal Learning and Cognitive Styles

One of the main assumptions regarding multimedia learning is that the "integration of graphics, video, sound, animation, and text can provide better ways of presenting information than any of these media can alone" [40]. In the educational context, the role of media is often restricted to the presentation of audio/visual materials that students passively watch in the classroom. Otherwise, the different media are usually presented as separate entities on the computer interface: for instance, text appears in one window while video clips and diagrams appear in other overlapping windows. This requires a continuous switching of attention on the part of users, so as to understand the relationship among the different

[1] http://infolab.stanford.edu/pub/voy/museum/pictures/display/0-3-XeroxAlto-text.htm

representations of content.

In order to increase the educational value of multimedia learning, it is essential to support teaching through designing innovative tools which ensure effective interactivity. To this end, it is essential to take into account how learners integrate information arising from different representations of the same and different information.

According to the *Cognitive Theory of Multimedia Learning* ([41], [42]), students engage in three important cognitive processes:

selecting the incoming verbal and visual information in order to represent a text and an image, respectively;

organizing the word and the image to create a verbally-based model and a visually-based model, respectively;

integrating, which occurs when the learner connects verbal and visual information.

Considering the above cognitive processes, Moreno and Mayer ([41], [42]) defined the following five principles on how to use multimedia in learning in order to help students achieve a meaningful knowledge.

1. The content to be taught has to be presented using two modes of representation – verbal and visual – rather than just one (multimedia effect);
2. the verbal and the visual information have to be provided at the same time (contiguity effect);
3. the verbal content has to be conveyed auditorily rather than visually – on-screen text – (split attention effect);
4. multimedia, contiguity, and split attention effects depend on individual differences in the learner; these effects are more important for low-knowledge than high-knowledge learners, and for high-spatial rather than low-spatial learners;
5. the multimedia content has to have few rather than many extraneous words and images (redundancy effect).

The foregoing multimedia principles are basic in design and development of successful instructional systems such as IMEs which aim at supporting efficient interaction of verbal and visual representations during learning activities.

The Cognitive Theory of Multimedia Learning is founded on the *Dual Coding* theory [43] by which cognition involves a verbal information processing system, specialized in dealing with language, and a visual information processing system, specialized in dealing with non-linguistic objects. According to this theory, the auditory content is directed to the verbal system, whereas animation is directed to the visual system. The Dual Coding theory has a main educational implication, which is contrasting with the emphasis on the dominance of the language experience in educational programmes in most Western countries: cognitive growth is effectively stimulated by a proportionate verbal and nonverbal experience. This means, for instance, that children who listen to the stories and look at the pictures recall more of the information than children who only listen [44]. Therefore, the design of technological tools for teaching, such as the IMEs documented in this thesis, need to consider this principle so as to implement effective learning. Fiaba Magica (Sect. 3.3.2), for example, according to the foregoing theories, provides the possibility to control the narrative animation which engages independent visual and auditory channels, expanding the working memory capacity.

It is particularly interesting to notice that both the Cognitive Theory of Multimedia Learning and the Dual Coding theory take into account that each student has individual learning peculiarities. Gardner [4; 5] argues that there is a wide range of cognitive abilities (Theory of Multiple Intelligences) and the purpose of education should be to help learners reach goals that are suitable to their range of intelligences. The Theory of Multiple Intelligences posits that there are multiple forms of knowledge, not only linguistic and logical-mathematical, but also interpersonal, intrapersonal, spatial, and bodily-kinaesthetic, that support and reinforce one another. Human beings indeed show different interaction modalities, ways of learning, acquiring, and maintaining knowledge and skills; this is further augmented by the impact of sociocultural factors. Active learners and visual learners are examples of students who thrive in contexts that extend beyond traditional teaching methods. Thus, the effectiveness of a learning activity is proportional to the number of intelligences addressed.

The introduction of IMEs at school pursues the above important goal: to consider the different cognitive styles, namely, the various forms of knowledge in order to answer pupils' actual needs. However, most computer-based educational activities offer only limited modes of engagement and rely on conventional mouse/keyboard/controller interfaces. These interfaces limit the naturally expressive abilities of students [45]. Hence the question: how can technology be used to foster student learning? The answer can be to design technological tools able to implement a high-level of interactivity and *multimodality*.

Multimodality occurs when different communication channels are used to convey or acquire information. Apart from this general definition, each scientific branch can describe the meaning of multimodality differently. In the context of the cognitive sciences, a modality is defined as a sensory modality, a perception via one of the perception channels (visual, auditive, tactile, etc.) [28]. In this case multimodal is synonym of multisensory.

From the HCI point of view, a modality is considered as the representational modality, which is a single means used to express information (spoken and written language, maps, images, etc.). A multimodal system supports communication with the user through different modalities such as voice, gesture, touch. The multisensory presentation of contents implies the:

1. selection of modalities; a careful consideration on the number and which modalities use;

2. mapping of modalities; to find natural relations between the chosen modalities and the content to convey;

3. combination, synchronization, and integration of modalities; to consider the spatial and temporal combinations of the sensory channel involved in interaction;

4. adaptation of multi-sensory presentation to accommodate changing tasks, context, and circumstances; to ensure flexibility of the system according to environmental changes and user skills [28].

By offering verbal and non-verbal information synchronized in time and space, IMEs, such as the Stanza Logo-Motoria (Chapter 3), offer a multimodal presentation of content that is very efficient, since it exploits visual and auditory channels. The Stanza Logo-Motoria, being a multimodal environment for learning, allows children to work together on interactive activities employing several perceptual channels in a physical environment. They are stimulated "to learn by doing" and to exploit their collaborative inclinations. Although the Stanza Logo-Motoria cannot replace teachers, it helps them motivate and reach more students simultaneously [46]. Instead of shaping the educational experience by means of technology, technology is tailored to a desired educational experience. This implies the reorganization of the curriculum around the intelligences, a educational shift from "teaching through instruction" to "learning by construction", the organization of teaching activity to allow multimodality, and finally engaging in more authentic methods of assessment than those used in traditional education.

2.2 Related Works

Given the theoretical framework described above, the research documented in this thesis is mainly focused on IMEs which are embedded within a physical environment (a classroom, a gym, a stage, a museum, etc.), augmenting the level of interaction and combining verbal (printed words, spoken words) and non-verbal (static graphics such as illustrations and dynamic graphics such as videos and animation) modes. These technological systems are able to make a physical ample space become an interactive and multimodal one. The physical approach of IMEs is also described by the term “blended reality” [47].
The focal point of the research is on the relationship between users (learners) and the IME in which they are located in order to learn. Moreover, the aim is also to verify if and how pupils’ understanding can be enhanced by adding nonverbal knowledge representations to verbal contents within an IME where words and pictures/animations are triggered by the learner’s actions during the learning session [48].
This work documents the development and validation of IMEs that have the aim of pursuing the following tasks:

- to draw attention to different aspects of the physical world;
- to experience a wide range of perceptual information;
- to create, explore, and discovery;
- to promote collaboration between children and others stimulating more verbal engagement and exchange of ideas [48].

A large field of research focuses on IMEs for learning in which audio/visual information is embedded into the physical world [19]. These systems use vision-based motion recognition techniques and non-invasive technologies, in order to analyze the movements of the user within a space. The computer-vision tracking and action-recognition techniques allow the pupil to naturally explore the augmented space without having to wear sensors. Thus, a room can provide for a real time multimedia feedback by exploring the physical environment [49].
The following are just a few examples of IMEs which, in recent years, have been developed to implement new modalities of learning in physical spaces. Since the late 1960s, researchers have been looking beyond the desktop, towards computer-augmented environments which exploit the advantages of physical spaces to augment and enhance the interaction with the real world itself. They do not replace the real world, on the contrary, they add to it [50].

In the field of education and entertainment, Krueger [51] explored the full-body interaction in a room-sized space. His IMEs, such as Videoplace (Fig. 2.1), have been used in the educational field as shared work spaces, scientific exploration for children, and physical therapy. Krueger created the Responsive Environment, a medium in which the "computer perceives a participant's actions and responds in real-time with visual and auditory displays" ([52] and [53]).

Figure 2.1: Videoplace, Myron Krueger - Full-body interaction in a room-sized space. Source: http://www.inventinginteractive.com

The ALIVE (Artificial Life Interactive Video Environment) project [54] improved on Krueger's system by means of computer-vision algorithms that can track the position and gestures of a person moving in front of a static background. ALIVE allows full-body interaction between a human participant and a rich graphical world inhabited by autonomous agents. This system can be used with interactive storytelling applications in which the user plays one of the characters in the story, whereas the artificial agents are all the other characters; furthermore, the autonomous agents can be modeled to act as a personal trainer, able to demonstrate how to perform and action, or a teacher.
The augmented physical space was further explored by Davenport and Friedlander [55] who developed Wheel of Life (Fig. 2.2), an interactive world situated in a real space; the space contained a narrative that could be actualized by the actions of the visitor. The room used light, sound, video, and computer displays. Each person in the space had an external human guide communicating with him/her via a computer. An important goal of this system was to teach students understanding of a "fabricated world" and to build communication skills which are fundamental for collaboration.
An important example of IME in physical space is the Music Atelier of "Città dei Bambini" (Children's City) at *Porto Antico* in Genova, Italy [56]. The Music Atelier (Fig. 2.3), consists of five games characterized by multimedia-multimodal interaction involving music, movement, dance, and computer

Figure 2.2: Wheel of Life - An interactive world situated in a real space. Source: http://www.flickr.com

animation. Children, by playing interactively with various devices and games, learn about physics, music, biology, and many other subjects. This active sensorized environment enables children in creating and modifying music through their full-body movement in order to learn basic music concepts and introduces them to artificial intelligence concepts.

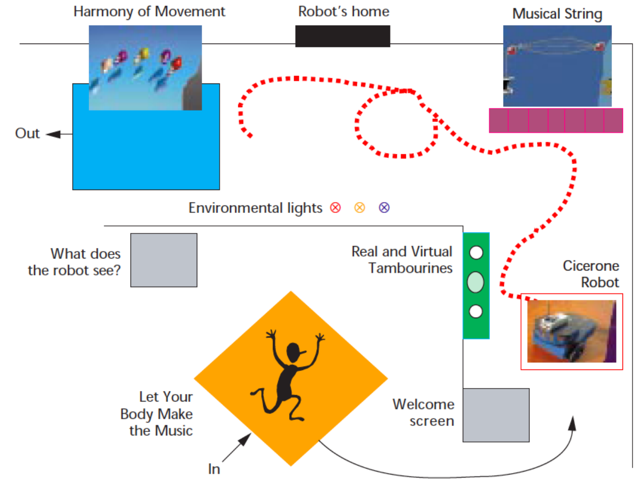

Figure 2.3: Music Atelier - Five games characterized by multimedia-multimodal interaction involving music, movement, dance, and computer animation. Source: [56]

KidsRoom [49] is a multimodal/multimedia and immersive storytelling environment which re-creates a child's bedroom (Fig. 2.4). Two of the bedroom walls resemble real walls, the other two are large video projection screens where images are back-projected. Six computers automatically control the room, the colored lights on the ceiling, and four loudspeakers; finally, four video cameras and one microphone are installed. The goal of this system is to stimulate a child's imagination in order to transform the room into a fantasy world. KidsRoom is an IME in which the computer interface is embedded within the physical world itself, the user does not require wearing special clothing and it is designed to allow multiple simultaneous users.

Figure 2.4: KidsRoom - An interactive and immersive storytelling environment. Source: http://vismod.media.mit.edu/vismod/demos/kidsroom/playsp.htm

Ambient Wood [57] is a system in which the physical world is a real wood. It involves the children experiencing the wood and wanting to learn more about the habitats they are exploring with their eyes and ears. Various tools and pervasive technologies provide contextually relevant digital information (Fig. 2.5).

Figure 2.5: Ambient Wood - An IME which presents multimedia digital information representing the relationship between organisms in the habitat. Source: [57]

The overall system is made up of various computer-based devices which provide for multimedia content coupled with the actual objects of the habitat. Ambient Wood presents digital information in the form of sounds, animations, videos, and images representing the relationship between organisms of the habitat.

The children are equipped with various technologies that enable them either to discover this information themselves (using PDA, Personal Data Assistant) or have them presented in a pervasive way, triggered by their physical presence, using Pingers (short-range radio), PDA, GPS (Global Positioning System), and tracking equipment. A key feature of Ambient Wood concerns the use of different channels to provide digital content for children, by moving around in an open space and through by using various devices. The overall goal of this project is to use "digital augmentation of physical activities" to extend learning: Ambient Wood implements a technology-mediated learning experience that helps children in integrating their understanding and knowledge through a process of reflection and action [57].
Very interesting to mention is the Hazard Room Game [58], implemented both in desktop and physical environments, that concentrates on teaching children about environmental health hazards (Fig. 2.6). The main goal of this project was to investigate, in direct comparison, the desktop and physical educational environments in the domain of young children's learning. The final results (obtained by means of qualitative and quantitative measurements) proved the advantages for physical environments over desktop environments.

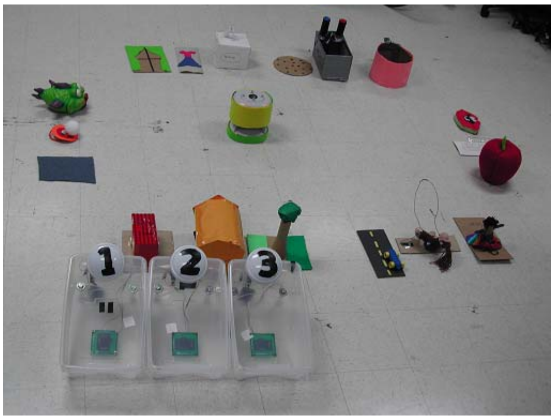

Figure 2.6: Hazard Room - The physical Hazard Room focused on teaching children about environmental health hazards. Source: [58]

A group of researchers at Arizona State University [59] developed SMALLab (Situated Multimedia Arts Learning Laboratory), an IME for high school teaching/learning that allows students to freely move in space while interacting with visual and auditory contents (Fig. 2.7). The use of SMALLab is based on the hypothesis that effective knowledge retention occurs when learning is embodied and multimodality is included during the act of learning.
The overall research project "Interactive and Multimodal Environments for Learning", documented in this PhD thesis and conducted by this author closely with the CSC of the University of Padova under the

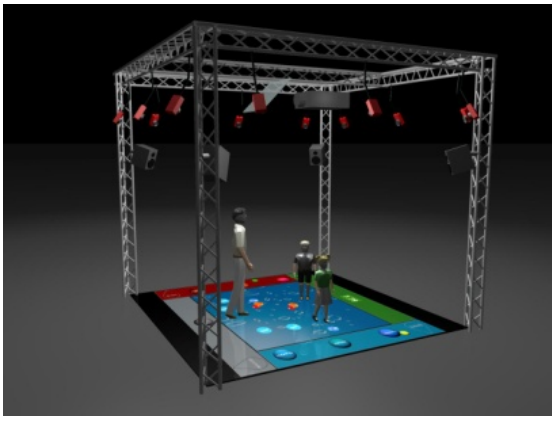

Figure 2.7: SMALLab - A Situated Multimedia Arts Learning Laboratory. Source: [59]

supervision of the Prof. Sergio Canazza Targon, involves bringing the potential benefit of IME to the every-day-teaching space, where it can be fully integrated with related educational resources. In particular, Chapter 3 describes the Stanza Logo-Motoria, an IME for learning, used as an alternative/additional tool for teaching in a real-world setting (educational institutions). Since 2009 the Stanza Logo-Motoria is permanently installed in a primary school where it is regularly used: to teach several school subjects, perform many interactive activities, and carry out an educational experimentation. In this way it was possible to verify that using such a system in a real school environment is complex due to a number of technical and practical variables that need to be take into account when IMEs are used the every-day-teaching activities.

For this reason the Stanza Logo-Motoria uses standard hardware and input/output conventional peripheral devices such as a webcam, two loudspeakers, and a video projector. The easy implementation of the system and the low cost of both the hardware and the software allow the use of the system on the part of public schools with real classes of schoolchildren. Moreover, the modular software architecture enables users to adapt the environment to different educational contexts, for example, learning a second language, studying History, or improving their spatial ability [60].

3

The Stanza Logo-Motoria: an Interactive and Multimodal Environment for Learning

3.1 Introduction

The Stanza Logo-Motoria is an Interactive and Multimodal Environment (IME) that this author conceived in 2008 with the idea to pursue a specific educational objective: to have a classroom transform into an Interactive and Multimodal Environment in order to provide schoolchildren with alternative learning methods that enable them to meet their real educational needs. At that time, this author, acting as a special needs educational teacher, had the task to support primary school children with learning difficulties and/or disabilities to optimize their potential within a variety of educational settings. The two-year experience of teaching a pupil with Cerebral Palsy brought to light the necessity on the part of the educational institution, to implement innovative instructional paths exploiting the motor component of learning. Technology can help reach this goal; the Interactive and Multimodal Environments, such as the Stanza Logo-Motoria, can be a good path to follow.

The Stanza Logo-Motoria indeed implements an innovative way of teaching, which offers greater access both to knowledge and interaction with others and the environment enabling teachers and schoolchildren to discover the enactive approach towards teaching/learning. This system helps teachers to deal with pupils with severe disabilities who need to learn by means of alternative methodologies and tools that often need to be tailor-designed for them.

Resonant Memory, the first application of the Stanza Logo-Motoria – based on the EyesWeb open software research platform[1] for the design and development of real-time multimodal systems and interfaces – was designed by this author and developed in 2008 in collaboration with the InfoMus Lab of DIST (Department of Communication, Computer, and System Science) of Genova University. In 2009, the Stanza Logo-Motoria was permanently installed at the "E. Frinta" Primary School of the "Gorizia 1" Comprehensive Institute of Gorizia (Italy) where it is still experimented with a large number of schoolchildren. The Resonant Memory application, utilized at school with experimentation purposes, was developed by CSC (*Centro di Sonologia Computazionale*) of Padova University [10].

Since 2009, both teachers and schoolchildren are using the system with interest and enthusiasm. Teachers believe the success of this system is due to its main features:

- the modular software architecture which enables teachers to adapt the environment to a variety of educational contexts;
- simple mapping strategies;
- the possibility to easily customize the multimedia contents.

Learners are still motivated by:

- the possibility to freely move in space activating sounds events (interactivity);
- the collaborative component of activities: there are a range of tasks that require a constant interchange with others/their peers and with multimedia contents using the modes of speech, image, movement, gesture and sound (multimodality);
- the opportunity to discover ever new and original multimedia contents (surprise effect).

The Stanza Logo-Motoria is an IME where "learning in movement" is the primary focus (Fig.3.1): by entering this physical environment and performing previously arranged gestures, the user "augments" the real experience with auditory and/or visual content. In order to cognitively stimulate children, the system provides video and audio contents, without replacing the real-world in which pupils are comfortable. The main goal is to augment the teaching activities by means of technology because it enhances the human capabilities [50]. Moreover, making educational activities fun, immersive, and engaging is another important goal of this research project.

[1] www.infomus.org

Figure 3.1: The Stanza Logo-Motoria at "E. Frinta" Primary School, Gorizia, Italy - Learning in movement.

3.2 The Stanza Logo-Motoria System Architecture

The Stanza Logo-Motoria is a hardware and software system, developed first by using the EyesWeb XMI platform[1], with which it is possible to create an IME: a real environment "augmented" with sounds and images by means of video-based tracking techniques. The Stanza Logo-Motoria consists of an ample empty space namely, a classroom, a gym, a stage, where the presence of a user and their movements trigger the playback of the audio/visual content. By means of the Stanza Logo-Motoria system, it is possible to subdivide the space into several areas; by entering the various areas and performing pre-set gestures, the user activates the auditory and/or visual content connected to them.
The Stanza Logo-Motoria is a modular system (Fig.3.2) which consists mainly in:

Input Components, standard sensors (e.g. a webcam) which track the user's body movements and gestures;

Real Time Processing Component, a software for multimodal interaction, which processes the input signal in order to calculate low-level motion features [61], is used to estimate a) how the user occupies the space and b) the quality of gestures the user performs; this information is the input of other software modules, which manage the interaction logic and the real-time rendering of audiovisual contents;

[1] www.infomus.org

Output Components, conventional devices such as loudspeakers and video projectors which allow the multimedia reproduction.

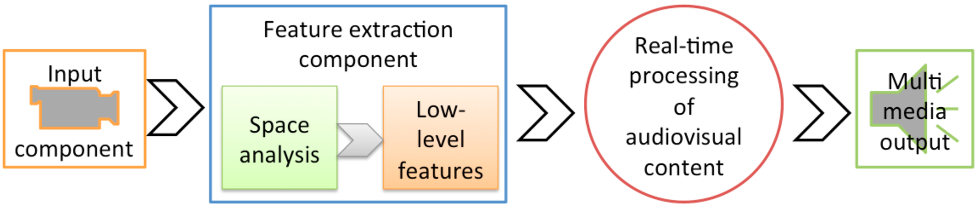

Figure 3.2: The overall system architecture of the Stanza Logo-Motoria - The Stanza Logo-Motoria is a modular system which consists mainly in Input Components, Real Time Processing Component, and Output Components.

This generic architecture can be used for specific educational objectives which are achieved by means of different applications:

1. the Resonant Memory application which can be used to teach several school subjects;
2. the Fiaba Magica application which may be useful to support a strengthening path of gestural intentionality of children with multi-disabilities;
3. the PlayAndLearn application which has the aim of teaching new vocabulary and meanings.

These applications are fully explained in the following paragraphs.

3.3 The Applications of the Stanza Logo-Motoria

The Stanza Logo-Motoria implements various interactive and multimodal activities depending to the software application used. At present, three different applications have been developed: Resonant Memory, mainly used to teach several school subjects in an alternative manner; Fiaba Magica, implemented to support learning for children with multiple-disabilities; and PlayAndLearn useful to teach new vocabulary and meanings.

3.3.1 Resonant Memory

Running the Resonant Memory application, the space, acquired by a webcam, is subdivided in nine areas: eight peripheral and one central (Fig. 3.3). The user's presence within a specific area triggers the audio reproduction of the corresponding audio content. Noises, music, and environmental sounds are

synchronized to each peripheral area, whereas the central area is synchronized with an audio reproduction of the contents to be taught (e.g. a story) that contain the elements to be connected with the sounds positioned in the various peripheral areas. The child listens to the content reproduced in the central area; reaches – by freely moving without using sensors – the different peripheral areas experimenting the sounds; and finally, enjoying the game, "composes the soundtrack" of the lesson [10; 11; 12].
In particular, with the Resonant Memory application the use of body movement associated with the sound widens the range of possibilities to access knowledge from an enactive point of view [37]. Teaching contents, by becoming "physical events", which occur around the child, and triggered by the children themselves, activating the motor aspect of knowledge [34].

Figure 3.3: The Resonant Memory application - This application subdivides the space in nine areas which are synchronized with the reproduction of audio content.

3.3.1.1 The Resonant Memory System Architecture

The first version of the Resonant Memory application is based on a single software module, developed in the EyesWeb XMI environment. EyesWeb XMI[1] is a software developing tool based on a graphical paradigm and applications which are developed connecting, by means of virtual cables, objects that implement basic processing tasks. The Resonant Memory EyesWeb patch (Fig. 3.4) analyzes the input video stream, the parameters of the interaction logic, and the audio rendering task (this application does not require video rendering).
In the input stage the signal from the webcam is processed in order to extract several low-level features – the trajectory of the centre of mass and the contraction index – related to the user's movements. The preliminary background subtraction is achieved via a statistical approach: the brightness/chromaticity distortion method [62]. The second stage controls the interaction logic: as a function of the features

[1] www.infomus.org

representing the user's movements, the application manages the transitions among four states: exploration, story, pause, and reset. Finally, the output stage controls the playback and mixing of a set of pre-recorded audio files [12].

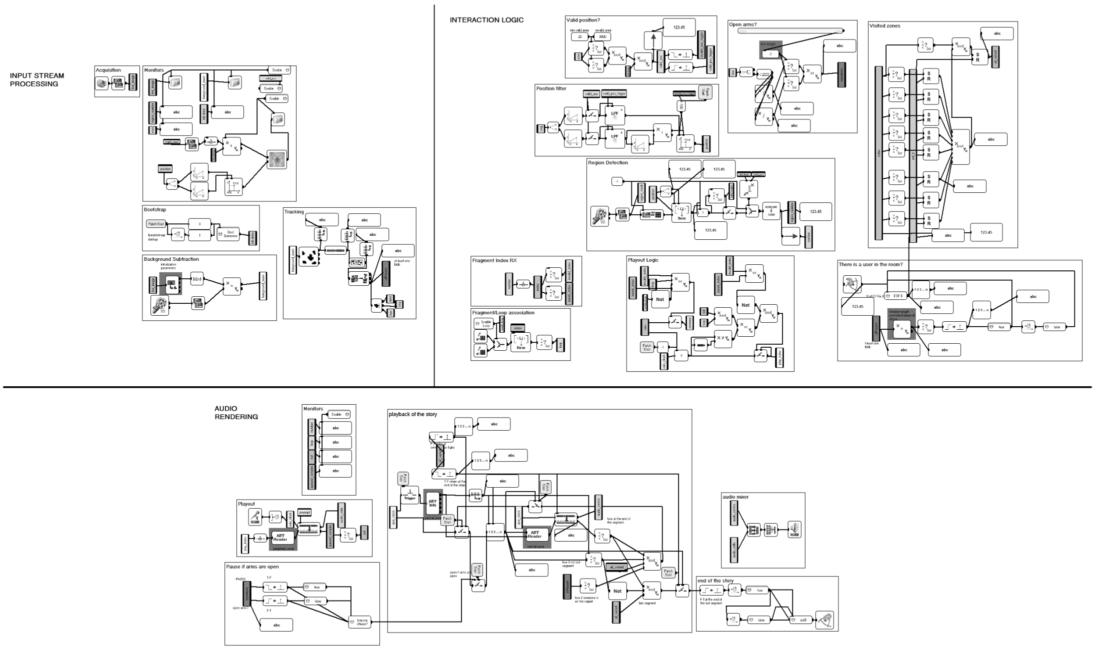

Figure 3.4: The Resonant Memory software module - The EyesWeb XMI software patch.

In 2012, a completely new software was developed by the CSC of the Padova University during the master's thesis work of Leonardo Amico, supervised by Sergio Canazza, Antonio Rodà, and Serena Zanolla [63]. This new software version of the Resonant Memory application has been devised with the goal of enabling teachers – with little or no computer competency – to easily manage educational activities by means of the Stanza Logo-Motoria and create their own multimedia contents by themselves. The software has been written in C++, using the Openframeworks libraries (an open-source tool-kit for interactive applications) and the OpenCV library for Computer Vision.

Two main modules compose the software: the TrackingEngine and the AppManager. The TrackingEngine module manages the video data stream acquisition and the motion feature extraction. The TrackingEngine module creates a statistical model of the background using a certain amount of static video frames. The video frames of the users moving in space are compared with the background model in real-time and a "confiance image" is produced. The confiance image is further filtered and processed and finally, a thresholding algorithm, on the base of the statistical information about the regions of the image, creates a binary map in order to identify the contours of the foreground figures (blob tracking).

Figure 3.5 shows the image processing performed by the software in order to obtain the user's silhouette.

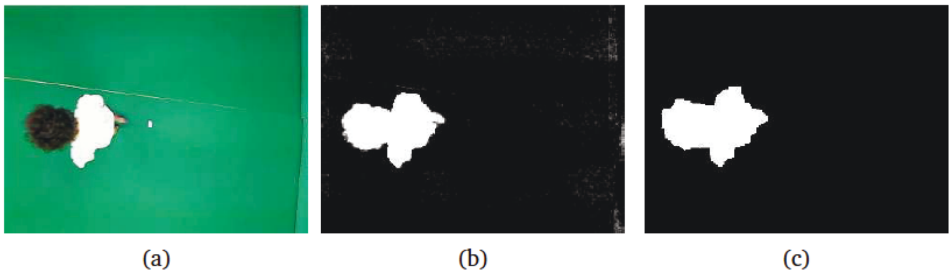

Figure 3.5: Image Processing - The image processing required to extract the user's motion features. (a) the video frame, (b) the confiance image, (c) the silhouette of the user. Source: [63]

On the base of this information, the software extracts the user's motion features and sends them to the AppManager module.

The AppManager module acts as a host for the different applications of the Stanza Logo-Motoria. In particular, the AppManager module:

- manages the auditory and visual feedback according to the interaction logic of each application;
- provides a graphical interface (Fig. 3.6) which allows teachers to monitor the users' movements and to choose among different applications and multimedia contents.

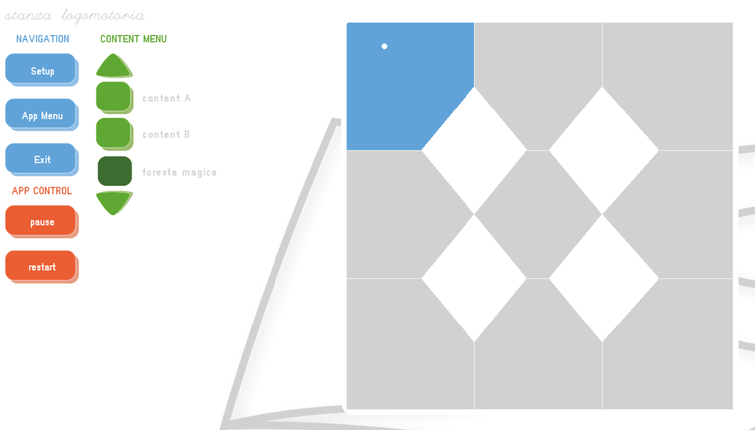

Figure 3.6: A screenshot of the graphical user interface of the Stanza Logo-Motoria in Resonant Memory modality - The figure shows the nine areas and the graphic white point 2D representing the user's position in the environment; on the left side there are several buttons useful to control the application (in orange), to navigate the system (in blue), and to choose the contents (in green). Source: [63]

3.3.1.2 Uses of the Resonant Memory Application in the Educational Context

The Stanza Logo-Motoria, by using standard hardware and simple strategies of mapping, aids both students and teachers by allowing innovative ways of learning and teaching. In fact, the use of the Stanza Logo-Motoria at school has demonstrated that by working with the teachers on the various subjects it is possible to develop many effective educational activities [60]:

1. As a tool to activate the written production of a story. In this case, the child is encouraged to invent a story starting by exploring the physical and auditory space;

2. To help study a school subject such as History. The teacher connects the content of the lesson with the central area, whereas the sound events, connected to the content, are synchronized with the peripheral areas. The child a) listens to the content reproduced in the central area, b) reaches the different peripheral areas experimenting with the sounds, and c) "fills the content of the lesson with sound events". This study mode (auditory) is particularly indicated for children with reading comprehension problems because, according with the Simple View of Reading Theory [64], they require to practice listening activities focused on the phonological structure of spoken words. Moreover, children with SpLD (Specific Learning Difficulties), who often find the school experience frustrating because school activities are mainly based on reading and writing, need to use "multisensory learning" [65];

3. As a tool to develop the spatial ability on the part of students with severe visual impairments. The Resonant Memory application might be used as a means to develop the orientation, navigation, and mobility skills;

4. As a method to improve the communication skills of foreign children who are learning Italian as their Second Language: various speech exercises might be planned where the speech sounds are not familiar to the child because they do not belong to the phonemic repertoire of their mother tongue;

5. As a listening tool for ESL (English as a Second Language) classes. The Resonant Memory application develops schoolchildren ability to deal with real-life listening where the input is very fast, and the listener is under pressure to extract the gist of what is being said.

In ESL classes (point number 5 of the list above), the Resonant Memory application is used as a listening tool: children explore the eight peripheral areas and memorize the spatial coordinates of sound events; then, entering the central area, they activate the audio file of a story in English. Children need to listen carefully in order to quickly understand the content and trigger the sound suggested by the story – by entering the peripheral areas – at the right time.

Sometimes, in one or more peripheral zones, several sequences of the story are included leaving sound gaps corresponding to nouns, adjectives or verbs that the child must fill by calling the right word aloud. These motor-auditory experiences are generated by the pupils to understand and learn the meaning of sounds, words, segments, and complete sentences of the new language [12]. On the basis of this activity (ESL lessons), a preliminary study has been designed and conducted, following a validation protocol, which is described in detail in Chapter 4.

3.3.2 Fiaba Magica

Fiaba Magica (Fairy Tale) is an application of the Stanza Logo-Motoria conceived by this author with the aim of improving the residual movement and the gesture intentionality of pupils with severe impairments enabling them to interact with a tale by doing simple and limited gestures [11]. This application allows the user to reconstruct sounds and images of a tale by performing a simple gesture such as raising their arms sideward and moving within a specific space (Fig. 3.7).

In particular, using Fiaba Magica a child enters a specific area of the Stanza Logo-Motoria and activates a) the audio reproduction of the first part of the story and b) the projection on the screen of the corresponding image (for example two characters). Once the audio reproduction of the first part of the story is played, by raising their arms sideward, the user can play with the characters projected, animating them. Moving to the next area, the child activates a) the projection of a new sequence of the story which includes another set of characters and b) the audio reproduction of the narrative sequence itself. The third advancement in the story is performed in the same way as the other two [12].

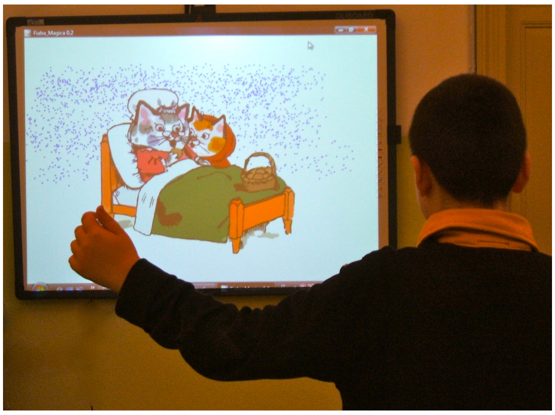

Figure 3.7: Fiaba Magica application - Fiaba Magica enables pupils to interact with a tale by raising their arms sideward and moving in space. Source: [11]

Children with multiple disabilities are often able to express communicative intentionality only by means of simple gestures and vocalizations, which can be enhanced and extended thanks to technology. Fiaba Magica is the opportunity to augment gestures with visual and sound stimuli bringing out all the intentional features. In general, Fiaba Magica can respond to a child's need to communicate, regardless their limited speech abilities associated with motor impairments. Since these children often find challenging, or even impossible, to walk without support (i.e. wheelchair), the Fiaba Magica application can cater for two users at a time or one user in a wheelchair accompanied by a helper [11].
However, at school, Fiaba Magica can be used by teachers with all the schoolchildren. Teachers indeed could convey spatial concepts, personal awareness of the different body parts, how the body moves in relation to other bodies and objects and the understanding of where the body is in an environment, and to learn new words and meanings.

3.3.2.1 The Fiaba Magica System Architecture

The Fiaba Magica application, developed by [11], uses only the first stage of the EyesWeb patch used for the Resonant Memory application, since the control of the interaction logic and the real-time rendering of the audio/video contents are being performed by an Adobe Air application that also provides a GUI for system configuration (Fig. 3.8). The communication between the EyesWeb patch and the Adobe Air module is achieved by means of the Open Sound Control (OSC) protocol [66] and a Java server, named Flosc[1], that acts as a gateway between OSC and Air, allowing messages to go in both directions.
An enhanced version of the Fiaba Magica application was developed in 2012 by CSC of the University of Padova during the master's thesis work of Marco Calabrese, supervised by Sergio Canazza, Antonio Rodà, Maja Roch, and Serena Zanolla [67].

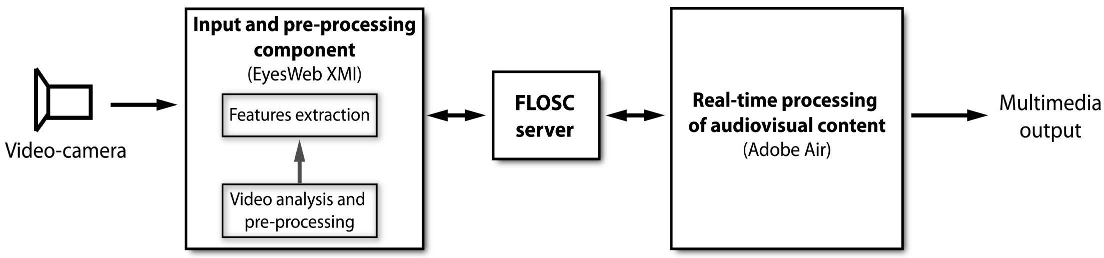

Figure 3.8: Fiaba Magica system architecture - The system architecture of the Fiaba Magica application. Source: [11]

[1] http://benchun.net/projects/flosc/

3.3.2.2 Preliminary Testing of the Fiaba Magica Application

In collaboration with parents and teachers, in February 2011, a preliminary usability test of the Fiaba Magica application was carried out; two pupils with severe impairments, who were only available for two days, were involved:

1. a 6-year-old child with left-sided hemiparesis who does not walk and uses a wheelchair;
2. a 12-year-old girl with cerebral palsy and lack of muscle coordination (ataxia).

The main aim of the two-day preliminary usability test was to observe how these representative users interact with Fiaba Magica: it was possible to notice that both pupils, by means of the Fiaba Magica:

- were motivated to move in space;
- improved their voluntary movement of their arms laterally which – in such a situation – is disturbed by synkinetic movements;
- enhanced their capability to hold their head up in order to follow the images with their eyes;
- improved their balance in standing and walking;
- increased their attention span;
- interiorized the temporal concepts (before, now, after).

Moreover, these users showed an exceptional degree of enthusiasm and enjoyment during the interactive listening experience. Due to physical and mental impairments, children with severe disabilities have few opportunities to engage in independent leisure activities which instead the Fiaba Magica application may provide. The Fiaba Magica application combines a comfortable and safe environment where the senses can be stimulated by the combination of sound and video events. Through this interactive experience users find their own ways of learning and communicating [12].

3.3.3 PlayAndLearn

Regarding the last educational activity mentioned in section 3.7 ("to learn new words and meanings"), a specific version of the Fiaba Magica application was developed by CSC of the University of Padova during the master's thesis work of Marco Calabrese, supervised by Sergio Canazza, Antonio Rodà, Maja Roch, and Serena Zanolla [67]: the PlayAndLearn application, an interactive and multimodal interface by which vision, hearing, and the body movements are used to trigger audio/visual content.

By means of PlayAndLearn, pupils easily acquire new words and meanings: the multimedia content is indeed constituted by the audio reproduction of spoken words and sentences and the visual reproduction of animations/images. To put it more clearly, the PlayAndLearn application works within the Stanza Logo-Motoria whose space is subdivided in three areas; each area is synchronized with a) the audio reproduction of a spoken word and its meaning and b) the video projection of two animations/images: one animation/image corresponds correctly to the meaning heard, whereas the other one corresponds to an extraneous word/object. The user's task is to choose the animation/image that correctly corresponds to the definition they have listened to. The choice is made by raising the right or left arm sideward according to the position of the image on the screen. The user's movements and gestures are captured by the Microsoft Kinect Sensor (Fig. 3.9).

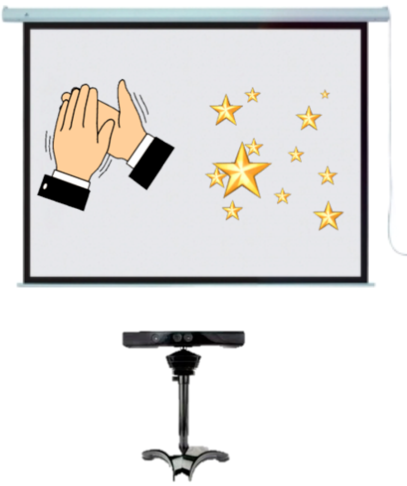

Figure 3.9: The PlayAndLearn application - The user chooses the correct animation by raising the right or the left arm sideward according to the position of the image on the screen.

3.3.3.1 The PlayAndLearn System Architecture

Figure 3.10 shows the overall architecture of the PlayAndLearn system. The Microsoft Kinect Sensor tracks the user's movements and gestures, whereas a specific software application manages the interaction logic and the real-time rendering of the audio/video contents.

The PlayAndLearn software is composed by two modules: the first module is an application, developed in C++, which: a) receives the data deriving from the Microsoft sensor, b) recognizes the back and forth movement of the user, c) tracks the gesture of widening the arm laterally, and, finally, d) via UDP (User Datagram Protocol), sends the OSC (Open Sound Control) packets (containing the data which describe the user's movements) to the second module.

The second module is an Adobe Flash application, developed in ActionScript 3, which receives the data about the user's movements and manages the audio/video contents according to the activity previously described.

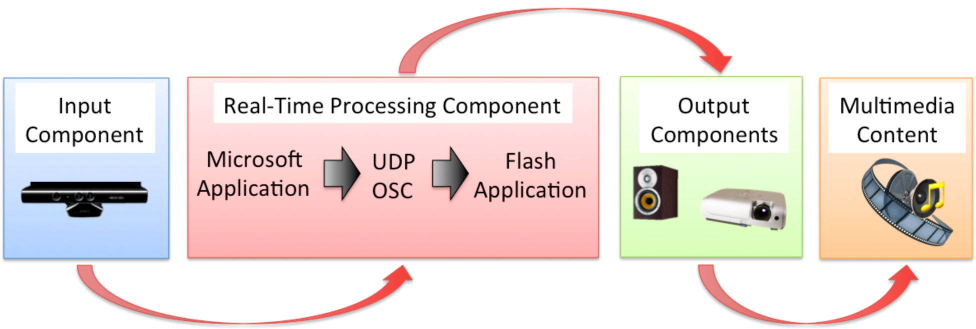

Figure 3.10: PlayAndLearn - The PlayAndLearn system architecture.

3.3.3.2 The Preliminary Study of PlayAndLearn

At the time of writing, a validation protocol of the PlayAndLearn application is being planned: a between-subjects experimental design will be applied to a First class at primary school level. This preliminary study will aim at verifying whether vocabulary acquisition obtained by means of the PlayAndLearn application is more effective than vocabulary acquisition obtained by a traditional PowerPoint presentation.

3.4 Devices

The setup (Fig. 3.11) of the Stanza Logo-Motoria installed at the "E. Frinta" Primary School of Gorizia ("Gorizia 1" Comprehensive Institute) and used for the preliminary study consists of an empty classroom measuring a 5x7 meters. A webcam (Logitech QuickCam Ultra Vision) is installed in the centre of the ceiling at a height of 4 meters; the webcam is connected to the computer by means of an active USB 2.0 signal booster cable 5 meters long. A personal computer (Acer Veriton M460 – Processor: Intel® Pentium® Dual CPU E2200 2.20 GHz – Memory (RAM): 1.00 GB – Operating System, 32 bit: Windows Vista Business) runs the Resonant Memory, Fiaba Magica, or PlayAndLearn application. The audio feedback is provided by an amplifier (GBC XPP 600 250W) and two loudspeakers (Sony, model: SS-L50) while the video feedback is given by means of a video projector (Sanyo PRO Wide Multiverse – model: PLC-WXL46A). The room is lit by diffused lighting in order to avoid the projection of shadows on the floor.

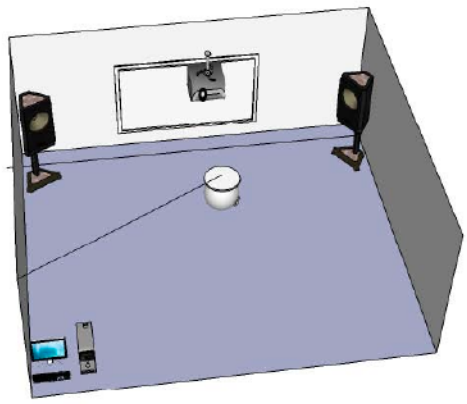

Figure 3.11: The Stanza Logo-Motoria setup - The basic hardware components of the Stanza Logo-Motoria: an empty space, a webcam on the ceiling, a computer, two loudspeakers, and a video projector. Source: [25]

4

The Preliminary Study of the Stanza Logo-Motoria in Resonant Memory Modality

4.1 Introduction

This Chapter addresses the research method utilized in this study – including the description of participants, material, data collection procedure, and data analyses – in order to determine whether the use of the Stanza Logo-Motoria as a listening tool in ESL (English as a Second Language) classes has the hypothesized effect on the participants.

In order to check the potentialities of the Stanza Logo-Motoria – in Resonant Memory modality – a "quasi-experimental preliminary study" was conducted [14]: the "pre- and post-test non-equivalent group design" – where "the experimental and control groups were not equated by randomization" [13] – was identified as the best educational research method.

The English as a Second Language (ESL) listening proficiency of two treatment groups – experimental and control group – were compared: two comparable classes of students, where the random selection of participants is impracticable because of the school organization. The experimental group only used the Stanza Logo-Motoria (experimental factor), whereas the control group used the language laboratory to perform the ESL listening activities.

For two school years (2011 and 2012), from February to June, the quasi-experimental study was conducted by this author at the "E. Frinta" Primary School ("Gorizia 1" Comprehensive Institute, Gorizia, Italy).

4.2 The Research Question

The purpose of this preliminary study was to try out and test the impact of implementing the use of the Stanza Logo-Motoria on ESL listening competences of two comparable classes of primary school children. The quasi-experimental study attempted to answer the following research question: *Is the primary school learners' proficiency in ESL listening comprehension improved when the Stanza Logo-Motoria – in Resonant Memory modality – is used as a listening tool?*

4.3 The Experimental Hypothesis

The main aim of this preliminary study was to verify whether pupils, who used the Stanza Logo-Motoria as a listening tool in ESL lessons, improved more significantly in word recognition and oral language comprehension compared to those performing traditional listening activities – by means of a CD player and two loudspeakers – in the language laboratory.
The use of the Stanza Logo-Motoria as a listening tool – in Resonant Memory modality – was the independent variable, whereas the proficiency in ESL listening comprehension was the dependent variable, which was observed to determine what effect the independent variable may have on it [13]. The proficiency in ESL listening was considered a valid indicator of the independent variable and was measured by means of pre- and post-assessment tests (Sect. 4.4.2.1).

4.4 Method

4.4.1 Participants

The participants involved in the preliminary study were two comparable classes:

- in 2011 (first experimental stage), two Third Classes (eight-year old children); an experimental group of 15 pupils (6 females, 9 males, 2 foreign students, 1 pupil with Opposition Defiant Disorder) and a control group of 14 pupils (11 males, 3 females, 4 foreign students); the application of the experimental factor was assigned randomly;

- in September 2012 (retention test) the same classes – considering exclusively the test-takers involved in the first experimental phase – a number of 12 pupils in the experimental group (9 males, 3 females, 4 foreign students) and a number of 13 pupils in the control group (8 males, 5 females, 1 foreign student);
- in 2012 (second experimental stage), the same classes (nine-year old children); the experimental group, a Fourth class of 16 pupils (12 males, 4 females, 7 foreign students) and the control group, a Fourth class of 12 pupils (9 males, 3 females, 2 foreign students). In this experimental phase, the application of the experimental factor was inverted.

Throughout the experimental period, the two treatment groups attended the ESL lessons separately in order to avoid contamination [13]. Moreover, school children attending private English courses or speaking English with parents at home were excluded from the participants of this study.

4.4.2 Material

This section describes in detail the material used for: a) the pre- and post-assessment tests and b) the ESL listening activities during the experimental treatments.

4.4.2.1 Testing Materials

The quasi-experimental preliminary study implied the comparison of the pre-treatment situation with the final performance of the participants, in order to check whether the hypothesis verified the forecast: if pupils using the Stanza Logo-Motoria make more progress in listening comprehension of ESL – thanks to the experimental factor (the innovation introduced) – than those who do not use it. With this aim in mind, for each experimental phase, a pre- and post-listening assessment test was administered to both classes.

The pre-tests were administered prior to the experimental treatments; the post-tests were administered after the experimental treatments. Moreover, in September 2011 a retention-test – with the same structure of the experimental pre- and post-tests – was administered to both treatment groups in order to measure the memory retention factor.

The assessment tests were performance-based and were administered in the paper-and-pencil mode; the assessment tests were developed by this author in collaboration with the English teacher who supervised the content, the structure, and the instructions and answering procedures of each test. The assessment tests entailed the following two tasks to accomplish.

Task 1 - Word comprehension; the first paper-and-pencil test presented 3 strips with 5 pictures each; the children had to choose the strip that exactly depicted the list of 5 words they heard (maximum score 1). An example of this task is shown in Fig. 4.1; students had to listen to an audio file featuring the following 5 words: "cereal, meat, giant, a fountain, a statue"; by choosing the second strip – that contains all the 5 words/pictures – they gained 1 point, whereas by choosing another strip, they gained 0 points.

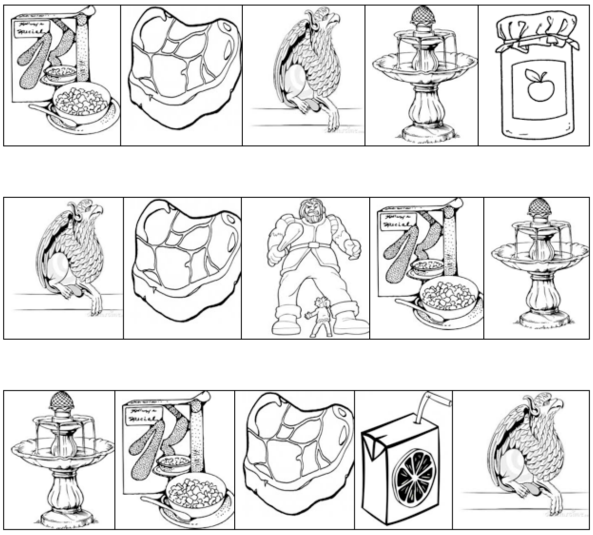

Figure 4.1: Word comprehension task - The first paper-and-pencil test presented 3 strips with 5 pictures each; test-takers had a) to listen to an audio file featuring a list of 5 words (e.g. "cereal, meat, giant, a fountain, a statue") three times and b) to choose the strip that exactly depicted the list of words they heard.

Task 2 - Sentence comprehension; the second paper-and-pencil test presented a grid of pictures depicting actions, situations, and contexts. Children had to listen to an audio file featuring a numbered (9) list of sentences three times. Then, they had to write the number of each sentence in the appropriate square (maximum score 9). Fig. 4.2 shows an example of the pictures that children had to match with the sentences heard. In this case, children listened to the following sentences:

1. "One! I have cereal and milk for breakfast and I have orange juice".
2. "Two! You can see a blue whale and dinosaurs".
3. "Three! What's the time Percy? It's two o'clock".
4. "Four! The leek and the daffodil are symbols of Wales".

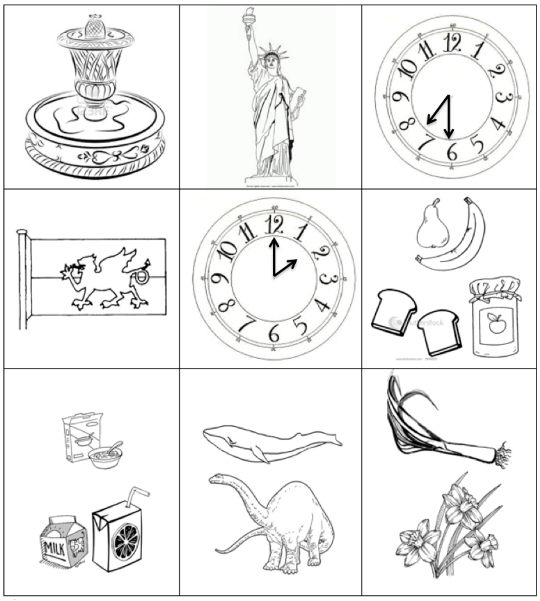

Figure 4.2: Sentence comprehension task - Pupils were presented with a paper-and-pencil test. They had to listen to an audio file featuring 9 sentences three times; then, children had to write the number of each sentence in the appropriate square.

5. “Five! I usually have toast, jam, and fruits for breakfast”.
6. “Six! There’s a famous statue in the centre of the city”.
7. “Seven! What time do you have breakfast, Nasareen? At half past seven”.
8. “Eight! The national flag of Wales is the red dragon”.
9. “Nine! There’s a fountain in the square”.

The pre-test contents were based on the linguistic skills acquired at school until that moment; the post-test contents were based on the linguistic skills acquired at school during the experimental treatment. The retention-test contents were based on the linguistic skills acquired at school during the first experimental phase. The content of the tests was the same for both the experimental groups. The audio material used for both the experimental treatment and the pre/post-assessment tests were those included in the syllabus and contained in the audio-CD which supplements the school books “Treetops Plus, Vol. 3 and 4” ([68] and [69]).

4.4.2.2 Experimental Material

This section describes in detail the audio material used throughout the experimental treatment by both the experimental and control group.

Audio material utilized for the first experimental phase. The audio files which supplement the schoolbook [68] were used; in particular, the audio files of the following lessons: *Identity cards*; *Numbers in town*; *Count to twenty*; *How old are you?*; *Cool cat* (Unit 1); *Shopping*; *Food and drink*; *I like fish*; *Do you like salad?*; *A new friend*; *Cool cat*; *At the shop*; *Money*; *Town mouse and country mouse* (Unit 2). A total of eight listening activities were carried out.

Audio material utilized for the second experimental phase. The audio files which supplement the schoolbook [69] were used; in particular, the audio files of the following lessons: *The Ring of Albion* (Unit 3); *Meal Times*; *Food*; *A terrible breakfast*; *What's the time?*; *At half past one*; *The Ring of Albion* (Unit 4); *Culture Scrapbook*; and *Around town* (Unit 5). A total of eight listening activities were carried out.

4.4.3 Procedure

4.4.3.1 Testing Procedure

Every year, in February, the pre-assessment test was administered close to the start of the experimental treatment – 1 or 2 weeks before – in order to avoid contamination of other variables (e.g. traditional ESL lessons). In May/June, the post-assessment test was administered close to the end of the experimental intervention (a week after). If the post-test is administered too soon, the long-term/delayed effect will be lost and only short-term measures are possible. With an excessive time lapse before the administration of the post-test, it becomes difficult to determine whether the effect is caused by the independent variable and not by other factors which have intervened since the experimental treatment [13].
The test administrator, in this case this author, provided test-takers with clear instructions on:

what to do for each task (described in detail in Sect. 4.4.2.1); test-takers had to listen to the audio files three times, look at a number of pictures on a worksheet, and choose the one that best described the listening; the children had to write down their answers;

what kind of response is required for each task; a cross for the word comprehension task and a number for the sentence comprehension task;

how and where to enter the response; close to the strip for task 1 and inside the squares for task 2.

The assessment tests were completed approximately in 30 minutes. The assessment was performed in the language laboratory by means of an audio system with two loudspeakers. Moreover, before the preliminary study, a pilot test was carried out in order to point out the faults of the test procedure [70].

4.4.3.2 Listening Procedure in the Language Laboratory

This section describes the procedure that the English teacher followed in conducting the ESL listening activities in the language laboratory during both the experimental phases.

Every week, the English teacher spent two hours teaching the control group in the language laboratory, where pupils remained at their desks and performed the listening activities by using an audio system with two loudspeakers. The control group completed the listening activities according to the instructions in their text books ([68] and [69]). The control group used the audio material contained in the audio CD which supplemented their text book. To support the listening activity the control group used the flashcards which supplemented the text book.

4.4.3.3 Listening Procedure in the Stanza Logo-Motoria

This section describes the procedure that the English teacher followed in conducting the ESL listening activities in the Stanza Logo-Motoria.

During both the experimental phases, every week, the English teacher spent two hours teaching the experimental group using the Stanza Logo-Motoria in Resonant Memory modality. As for the control group, the experimental group used the audio material contained in the audio CD that supplemented the text book.

All the participants carried out the ESL listening activities in the Stanza Logo-Motoria by using the Resonant Memory application. In pairs, they performed: a) the exploration of the eight peripheral areas containing the sounds (music, noises, environmental sounds but also activities such as “Listen and point” or “Repeat”); b) the listening task of the lesson – a story, a dialogue, a description – by entering the central area; and, at the same time, c) adding the sounds by reaching the correct peripheral area).

While a pair of participants was working together in the Stanza Logo-Motoria, the rest of the group was observing their performance outside the interactive space. The experimental group used the flashcards which supplemented the text book. In this case, the flashcards were placed on the wall near the area containing the audio file connected to them.

4.4.4 Data Collection

The preliminary study lasted for four months for two consecutive school years (2010/11 and 2011/12); this allowed the dependent variable to be tested for an appropriate period of time in order to acquire a sufficient amount of data to analyze.

The data gathered by means of the assessment tests were the number of correct answers for each participant. For every correct answer, test-takers gained one point; the maximum score obtainable (Task 1 +

Task 2) was 10, which means that participants answered 10 items correctly. The data obtained from the assessment tests were firstly subdivided into two categories – the experimental group and control group – and then into the following three and two categories:

- the number pupils 1) who increased the level of performance in ESL listening comprehension; 2) those who decreased; and 3) those who maintained the initial level;
- the number of pupils 1) who increased/maintained the level of performance and those 2) who decreased the initial level.

The Fisher's exact test and the Effect size test were the statistical tests used to assist data processing and analysis.

The Fisher's exact test was identified as the appropriate nonparametric statistical test for this preliminary study a) thanks to its utility for small samples (less than 10 in any cell of the data table); b) since it does not make any assumptions about distributions of scores (distribution-free); and c) because it is used in the analyses of nominal data. The Fisher's exact test was implemented to examine the significance of the contingency between the two treatment groups by measuring the deviation from the null hypothesis which means that there is no relationship between the two treatment groups and the experimental factor has no effect. The Fisher's exact test was used to determine whether there were non-random associations between the two categorical variables (the two different listening methods).

The Effect size is a standard measure of the degree to which a phenomenon is present or the degree to which a null hypothesis is not supported. The Effect size is useful to quantify the difference between two groups; it is the measure of the effectiveness of the treatment, in this case, the use of the Stanza Logo-Motoria. The Effect size shows "how big the effect is, something that the p value – statistical significance – does not do" [13].

4.4.4.1 Results of the First Experimental Phase

Before the first experimental treatment (pre-assessment test, February 2011), the experimental group (15 pupils) and the control group (14 pupils) both presented the same degree of word recognition and oral language comprehension: the average group score was, respectively, 7.9 ($\sigma = 3.27$) and 8 ($\sigma = 3.36$); this means that both treatment groups started at the same level of oral language comprehension competency.

Following the first experimental treatment (post-assessment test, June 2011), the average group score differed: the average experimental-group score was 9.1 ($\sigma = 1.43$), the average control-group score was 8.07 ($\sigma = 2.21$); the average experimental-group score increased compared to that of the control group. In order to verify whether the difference between the data obtained was solely due to changes introduced by the independent variable (the use of the Stanza Logo-Motoria) and not by random factors, the Fisher's exact test was applied on the values shown in the tables 4.3 and 4.4.

Fisher test 2x3	Increase	Decrease	No-Change	tot	P VALUE
Exp. Group	6	1	8	15	p=0,074
Cont. Group	4	6	4	14	> 0,05

Figure 4.3: Fisher's exact test based on 2 x 3 contingency tables - This test was applied to: the number of pupils who increased the level of performance, those who decreased, and those who maintained the initial level.

Fisher test 2x2	Increase/ No-Change	Decrease	Tot	P VALUE
Exp. Group	14	1	15	p= 0,0352
Cont. Group	8	6	14	< 0,05

Figure 4.4: Fisher's exact test based on 2 x 2 contingency tables - This test was applied to the number of pupils who increased/maintained the level of performance and those who decreased the initial level.

The results of both the pre- and post-tests showed a bunching of scores at the upper level (in particular, 8/15 maximum scores in the experimental group). Due to the high number of pupils who achieved the maximum score (ceiling effect), the Fisher's exact test result was $p > .05$ (not statistically significant). The overly easy content of the tests probably compromised the assessment of the participants' real listening skills.

However, the Fisher's exact test based on 2 x 2 contingency tables (Fig. 4.4) confirmed that the difference between the two groups was statistically significant ($p < .05$). Moreover, the Effect size test applied to post-tests results ($d = .55$) and to pre-/post-tests results of the experimental group ($d = .47$) showed that the effectiveness of the experimental factor was *moderate*.

Despite the small sample size and the ceiling effects which occurred, the results above indicate a moderate but statistically significant advantage of using the Stanza Logo-Motoria – in Resonant Memory modality – over the language laboratory.

4.4.4.2 Results of the Retention Test

In September 2011, a retention-test was administered to both treatment groups in order to measure the memory retention factor. The retention-test content was based on the linguistic skills acquired during the first experimental phase. Since the composition of the two treatment groups had changed, the test-takers involved in the first experimental phase were exclusively considered: a number of 13 pupils in the experimental group and a number of 12 pupils in the control group.

III B	Post-Test	Retention-test
Control Group	03/06/11	20/09/11
1 AB m	10	10
2 GB m	10	10
3 EB mfo	5	0
4 MC m	6	8
5 PF m	10	0
6 FG f	10	10
7 AG m	9	10
8 LK ffo	3	2
9 EK mfo	9	8
10 AL mfo	7	6
11 IL f	6	10
12 GR m	10	10
Average	7.92	7
Standard Deviation	2.32	3.87

III A	Post-Test	Retention-test
Experimental Group	03/06/11	20/09/11
1 GL m	8	8
2 RL m	10	10
3 FM m	10	8
4 IM f	10	10
5 EP m	10	10
6 MS ffo	10	10
7 CS f	8	3
8 RS m	6	9
9 NT m	10	10
10 EV f	10	10
11 JZ m	10	10
12 LZ m	10	10
13 MZ f	10	10
Average	9.38	9.08
Standard Deviation	1.21	1.89

Legend:

m = male
f = female
fo = foreign student

Figure 4.5: The post- and retention-tests results of each participant - In yellow test-takers who increased the level of performance, in red those who decreased, and in blu those who maintained the initial level.

Comparing the post-tests scores with the retention-test scores (Fig. 4.5), it was observed that: in the post-test the average experimental-group score was 9.38 ($\sigma = 1.21$), whereas in the retention-test the average experimental-group score was 9.08 ($\sigma = 1.89$). In the post-test the average control-group score was 7.92 ($\sigma = 2.32$), whereas in the retention-test the average control-group score was 7 ($\sigma = 3.87$). In order to verify whether the difference between the data obtained was statistically significant, the Fisher's exact test was applied to the following values [12]:

- the number of pupils who increased the level of performance (1 in the experimental group and 3 in the control group); those who decreased (2 in the experimental group and 5 in the control group); and those who maintained the initial level (10 in the experimental group and 4 in the control group); due to the high number of pupils (in particular, 9/13 in the experimental group) who achieved the maximum score in the retention test (ceiling effect), the Fisher's exact test result was $p > .05$ (not statistically significant); the overly easy content of the retention-test probably compromised the assessment of the participants' real listening skills;

- the number of pupils who increased/maintained the level of performance (11 in the experimental group and 7 in the control group) and those who decreased the initial level (2 in the experimental group and 5 in the control group); this test showed once again that the difference between the two groups was not statistically significant: $p > .05$.

On the contrary, the Effect size test applied to the post-tests scores suggested that, at the end of the first phase of the experimentation, the mean difference between the two groups was *moderate* ($d = .63$); this was due to a four-month use of different listening methods. Moreover, the Effect size test, with which the retention-test scores of each group were compared, suggested that the mean difference of the two groups was *moderate/large* ($d = .71$); this meant that the effectiveness of the experimental factor lasted for three months.

4.4.4.3 Results of the Second Experimental Phase

In June 2012, the second experimental stage – started in February 2012 – was completed. The tests were administered to the same school classes (nine-year old school children) but the application of the experimental factor was inverted. In this way, a Fourth class of 16 pupils used the Stanza Logo-Motoria (experimental group) and a Fourth class of 12 pupils (control group) used the language laboratory.
It is important to note that the composition of the classes changed in terms of total number and pupils: 3 new entries and 2 transfers in the experimental group; 2 transfers in the control group.
Before the second experimental treatment (pre-assessment test, February 2012), the experimental group and the control group presented different levels of word recognition and oral ESL comprehension competency: respectively, 6.56 and 7.17 (average group scores), $\sigma = 3.32$ and 2.19 (Fig. 4.6). Even if the competency level of the control group was not much higher ($d = 0.21$, small) than the competency level of the experimental group and both groups performed the traditional listening activities in the language laboratory for five months, the effectiveness of the Stanza Logo-Motoria had lasted for seven months.
After the second experimental treatment, the average group scores differed: the average experimental-group score was 8.88 ($\sigma = 1.83$) and the average control-group score was 7.42 ($\sigma = 2.36$); for the second time, the average experimental group score increased compared to that of the control group. In

II EXPERIMENTAL PHASE

	EXPERIMENTAL GROUP (16)		CONTROL GROUP (12)	
	PRE-TEST	POST-TEST	PRE-TEST	POST-TEST
Mean	6.56	8.88	7.17	7.42
SD	3.32	1.83	2.19	2.36

EFFECT SIZE TEST	
Pre-test experimental group/ Pre-test control group	*d* = .21 (small)
Post-test experimental group/post-test control group	*d* = .73 (moderate)
Pre-test/Post-test experimental group	*d* = .90 (large)

Figure 4.6: Second experimental phase - The Effect size test is useful to measure the effectiveness of a treatment, in this case, the use of the Stanza Logo-Motoria.

order to verify whether the difference between the data obtained was statistically significant, the Fisher's exact test was applied to the following values [12]:

- the number of pupils who increased their level of performance (14 in the experimental group and 7 in the control group), those who decreased (0 in the experimental group and 4 in the control group), and those who maintained the initial level (2 in the experimental group and 1 in the control group);
- the number of pupils who increased/maintained the level of performance (16 in the experimental group and 8 in the control group) and those who decreased the initial level (0 in the experimental group and 4 in the control group).

The Fisher's exact tests based on 2 x 2 ($p < .05$) and 2 x 3 ($p = .05$) contingency tables confirmed that the difference between the two groups was statistically significant. The Effect size test applied to post-tests results ($d = .73$) and to pre-/post-tests results of experimental group ($d = .90$) demonstrated that the effectiveness of the experimental factor was, respectively, *moderate* and *large*.
Despite the small sample size, results of the second experimental stage – indicating advantages of the Stanza Logo-Motoria, in Resonant Memory modality, over the language laboratory – imply that there is an overall advantage in using this system.

For the next school year, this experimental protocol could be implemented in another primary school where the Stanza Logo-Motoria system has been installed in order to verify if in a different educational context – with other pupils and teachers – the same results will be obtained.

4.4.5 Discussion on the Results

The results of the two-year preliminary study confirmed, to a certain extent, the experimental hypothesis: by means of the Stanza Logo-Motoria – in Resonant Memory modality – learners are more likely to comprehend and remember what they listen to, compared to pupils using the language laboratory. The ceiling effects, that occurred in the assessment tests of both the first experimental phase and the retention phase, prevented the assessment of the participants' real ESL listening competences. This was due to the to the overly easy content of the assessment tests.

More care was taken in the construction of the assessment test in the second experimental phase in order to avoid making the test too easy to complete. In this case, in fact, the results showed clearly that the experimental group increased the level of competency much more than the control group even though the application of the experimental factor (the Stanza Logo-Motoria) was inverted. The statistical analyses applied, confirmed that the differences between the experimental group and the control group were statistically significant: the difference between the data obtained was solely due to changes introduced by the independent variable (use of the Stanza Logo-Motoria) and not by random factors.

Despite the fact that the listening contents of the pre- and post-tests were the same for both groups, many subjects of the control group actually worsened. This occurred probably because:

1. they were more likely to be distracted from the listening: in fact, they performed passive listening activities sitting at a desk in the language laboratory;

2. they did not have, unlike the experimental group, the possibility to listen by means of a motor experience in a physical interactive space.

It is important to note also that a more regulated procedure has to be taken during the listening activities, especially for those of the control group: in the educational context many factors might influence the participants' attention and concentration, such as frequent interruptions of the lesson. From this point of view, the Stanza Logo-Motoria is effectively a more controlled environment.

On the basis of the results above, it is possible to say that, by means of the Stanza Logo-Motoria, schoolchildren may shift from passive to active listening and thus improve the ESL oral comprehension skills; meanwhile, teachers can supply them with an appealing educational setting, where sounds, noises, segments of spoken language, and complete sentences are used to enhance pupils' motor-auditory experiences. Words and sentences used to perform a task become part of the pupils' language knowledge, so speaking English becomes spontaneous, fun, and natural.

The Stanza Logo-Motoria facilitates attention and high concentration which, associated to motor experience in space, ensures improvement in ESL listening proficiency: the arrangement of sounds in space captivates the pupils' attention and allows them to spatially organize the acquired knowledge. Even after some time (memory retention test), children are able to recall the exact contents learnt during previous listening sessions and to identify the position of sounds in the interactive space.

The analyses of the video recording, together with the teachers, proved that this method of teaching increases the motivation to listen and consequently to learn new sounds, words, and phrases. In the teachers' opinion, learners are stimulated by a) the possibility to freely move in space activating audio-visual events (interactivity); b) the collaborative component of activities: there are a range of tasks that require a constant interchange with others and with multimedia contents using the modes of speech, image, movement, gesture, and sound (multimodality); and c) the opportunity to discover ever new and original multimedia contents (the surprise effect).

5

SoundingARM and Memory Sonoro: IMEs for Orientation and Mobility Skills Acquisition

5.1 Introduction

The inclusion of Interactive and Multimodal Environments (IMEs) within the range of aid tools for people with special needs has in the last decade sparked significant interest on the part of the research community. Many European research projects indeed, such as CARESS[1] (Creating Aesthetically Resonant EnvironmentS In Sound), Twi-aysi[2] (acronym for "The World Is As You See It"), and CARE HERE[3] (Creating Aesthetically Resonant Environments for the Handicapped, Elderly and REhabilitation), have pursued the development of IMEs that are able motivate people with disabilities to enhance creativity, imagination, communication, movement, and cognitive skills.

This main stream derives from the fact that IMEs provide innovative methods of interacting with multimedia contents, the environment, and others. In these multimodal "resonant" spaces, technology is used to "augment" the motor experience within the environment and the user-interfaces are not a mere drop down menus, a mouse, and different windows but rather full-body movements, gestures, and speech

[1] http://www.bris.ac.uk/caress/
[2] http://www.bristol.ac.uk/Twi-aysi/
[3] http://www.bris.ac.uk/carehere/

[71].

There is an ever-expanding research corpus that focuses on the implementation of IMEs for the rehabilitation of people with cognitive and/or sensory disabilities [72]. These technological environments can vary greatly in terms of design and content: from a more traditional concept which includes using lighting effects, colors, sounds, and so on [73; 74; 10], to virtual reality environments [75] and other technological spaces that use edge multimedia technology [76; 77].

Following this approach, SoundingARM [15] and Memory Sonoro [16], IMEs tailored for people with visual impairments, were designed, developed, and evaluated during the PhD period. The pursued aim of these IMEs is to supply users a visual and/or acoustic feedback according to the analysis of their gestures and speech. These IMEs are indeed able to generate multimedia feedback that supports sensory compensation, the evaluation process on the part of the operator (teacher, therapist), and the acquisition of specific skills. Within this type of environments, the users' actions are the point of focus, enabling the participant to engage with the content that the system provides.

5.2 SoundingARM: Assisted Representation of a Map

Assuming that the support of appropriate spatial information by means of sensory substitution may contribute to blind people's spatial performance, the SoundingARM (Assisted Representation of a Map) system [15] has been developed by the *Centro di Sonologia Computazionale* of the Information Engineering Department of the Padova University within the project "*Evoluzione dei sistemi di feedback acustico per la prima domiciliazione dei soggetti con inabilità visiva acquisita*" (Evolution of acoustic feedback systems for the homecoming of subjects with acquired vision impairment) funded by Friuli Venezia Giulia Region[1] (Italy) and carried out at the Regional Institute "Rittmeyer" for the Blind (in Trieste, Italy).

This author actively contributed to this project: a) by establishing and maintaining relationships with the Regional Institute "Rittmeyer" for the Blind and with mobility and orientation experts in order to identify the guidelines for the design and the development of SoundingARM and b) by conducting a preliminary usability test.

SoundingARM is a pointing-based system which aims at allowing users with severe visual impairment to build a spatial mental map of an unknown indoor environment. By means of this system the user can promptly explore a room by standing at the threshold and performing a simple gesture (finger pointing). By means of SoundingARM, users with severe visual impairment need to carry out two tasks: enter a

[1] Decree n. 1265/AREF dd. November 25th, 2010

room and, standing in the doorway, point in a direction in space with their arm. The system provides vocal information about the objects placed in that direction. These simple tasks aid blind people: (1) to immediately recognize the type of environment, (2) to quickly detect a specific object, or (3) to safely move in space (Fig. 5.1). Without having to wear sensors or hold pointing devices, the user receives

Figure 5.1: SoundingARM - Pointing in a direction, the system provides vocal information about the environment. Source: [15]

the main information about the type of environment (kitchen, bedroom, living room); the user does not need to perform a detailed tactile activity or a locomotor exploration and does not have to learn how to manage any kind of device.

If the users enter, for instance, a shop and at the entrance they point in a specific direction, the system provides vocal information about the type of products placed on the indicated shelves, the fixed obstacles that they will find on their way, and how to overcome them (in this case the users have to wear headphones).

Once the information above is obtained, the user navigates with more confidence in the environment or heads straight in the right direction, saving time.

The SoundingARM system aims at offering users with an auditory "first sight" of space, an acoustic map of space, allowing them to immediately be aware of the environment type and of the objects in it.

The SoundingARM system provides the user a means to explore a room by standing at the threshold

and performing a simple gesture: finger pointing. When the user points his/her arm around to find an object, the system promptly answers giving an audio feedback. Moreover, the user obtains the information regarding the position of an object through the gesture itself: the object is located in the direction in which the finger points.
Moreover, SoundingARM allows the user to experience the environment both from an exocentric overview (e.g. a map) and an egocentric (user-based perspective) "view" [78]. The system indeed, by using the blind users' current position data, given by the Microsoft Kinect sensor, can help them to orient themselves within the space by extracting information from the auditory map.

5.2.1 Theoretical Framework and Related Works

The development of orientation and mobility skills is basic for everybody in order to construct the spatial map of an environment. Since vision gathers most of the spatial information, people with severe visual impairment face difficulties in: a) moving in medium-scale spaces, where locomotion is needed for exploration; b) immediately recognizing the indoor environment type; or c) rapidly finding an object. The ability to move autonomously within space, to immediately recognize a place, to plan a route, or to choose the shortest path to take are all very important spatial competences in the everyday life of human beings [15]. Depending on their visual perception, these spatial behaviors are essential for everyone, especially for people with visual impairments. In these people, spatial knowledge translates into a mental representation which normally is acquired through supports, such as tactile maps, or creating a cognitive map based on the direct experience of the environment (locomotor and tactile exploration) [79].
The population with visual impairments can be subdivided into two groups, which is to say, those with low-vision and those who are blind. In order to obtain information regarding the environment, people with severe vision impairments need to use the other perceptual modalities (sensory substitution). In particular, those who are totally blind use hearing, touch, smell in order to perceive, interact, and move about their environments [80; 81]. The ability to independently orient, move, and navigate within space requires specific information-processing strategies.
In particular, *Orientation* refers to the awareness of one's position in space. It is a process of utilizing sensory compensation in establishing one's position and relationship to the objects of the environment [82]. Orientation depends upon the decoding of sensory information, which may be visual, auditory, kinesthetic, tactile, thermal, and/or olfactory. Blind people need to learn how to anticipate the regularities of the environment.
Mobility refers to the ability to move safely within the physical environment. Mobility is the "capacity, facility and the readiness to move" [82]. A visually impaired person needs to deal with different types

of obstacles (e.g. cars parked, potholes in the road, road signs, or road work).

Navigation refers to the process involved in traveling from one place to another, using mobility skills and orientation in the environment according to the desired route. The visually impaired people constantly need to update their orientation and position using the following three methods classified on the basis of kinematic order [83]:

- position-based navigation which relies on external signals (e.g. landmarks);
- velocity-based navigation which depends on external and proprioceptive signals;
- acceleration-based navigation which involves the user's linear and rotary accelerations (no external signals are required).

With respect to these three methods, visually impaired people are likely to have problems when navigating in unfamiliar routes since they lack visual information that is usually useful to make detours around obstacles and to have information regarding farther landmarks [83]. This is the reason why most of way-finding technological efforts to aid the visually impaired are limited to devices which can help people avoid the obstacles, namely the Long Cane considered the first obstacle detector. Currently, technological travel aids for visually impaired persons are contributing to create interactive multimodal spaces which, generally, aim at conveying auditory and tactile information to allow the user to reconstruct the spatial map of the environment. Vision substitution devices can be categorized into primary and secondary aids based on whether the device is used on its own or to supplement another device [84]:

- primary aids are used to deliver safe mobility (e.g. the long cane);
- secondary aids are used to supplement the primary device (e.g. the hand-held ultrasonic device for obstacle detection is used to augment the use of a long cane).

Travel devices can be subdivided in other categories according to their functionality: obstacle avoidance, route finding, environmental access assistive technology, and object finding. Technological travel aids can be also categorized according to; a) the technology used to obtain environmental information (ultrasonic, infrared, video signal, or GPS); or b) the modality through which the information is provided to the user: tactile (vibration), speech, sounds, musical tones; finally, c) vision substitution devices can be subdivided according to how the device is carried by the user: hand-held, in-pocket, or in-backpack [84].

The first electronic mobility aid was devised in 1897 (the Noiszewski's Elektrftalm); then – through the 1950s and 1960s – infrared and ultrasound, specifically the Electronic Cane, Mowat Sensor, and the

Figure 5.2: The Pathsounder - It is a mobility aid to assist the blind user to detect obstacles which his/her cane cannot reach. Source: http://webmuseum.mit.edu

Pathsounder (Fig. 5.2), were used to obtain environmental information [84].

Throughout the 1960s and 1970s, laser canes were produced to allow the obstacle detection. In the last decade or so, instead, interest in devices to support orientation and spatial sensing has increased. An example is Taking Signs[1] an infrared system which emits an audible warning when the user moves in proximity of the obstacle (Fig. 5.3).

There are many other technologies and applications that use ultrasonic or infrared systems [85; 86], GPS (Global Positioning System) technology[2], mobile technologies [87], radio frequency identifier (RFID) tags [88], or computer vision [89] to determine the traveller's current location.

To date, assistive technology systems for mobility, orientation, and navigation have been developed independently from one another, therefore, users often need for different devices to face the diverse phases of a travel; moreover, some of these devices are quite cumbersome, but nowadays, advances in signal processing, ICTs (Information and Communication Technologies), and devices miniaturization have made small size multi-purpose assistive technology systems a concrete possibility.

[1] http://www.talkingsigns.com/tksinfo.shtml

[2] http://www.csun.edu/cod/conf/2003/proceedings/140.htm

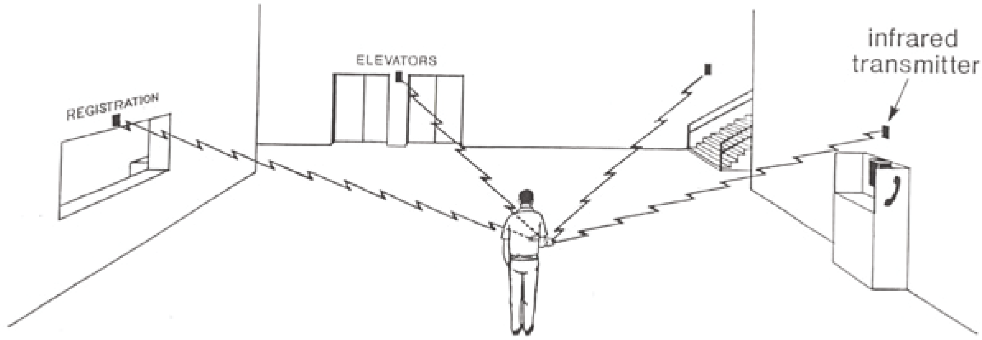

Figure 5.3: Talking Signs - The Talking Signs system employs infrared transmitters which provide users with auditory information via a hand held receiver. Source: http://www.ski.org

5.2.2 Objectives of the System

SoundingARM – Assisted Representation of a Map – was developed in order to help blind users to construct the spatial map of an environment; it is a primary technological travel aid for environmental access and object finding employing an infrared device and supplying the user with auditory information. SoundingARM, based on recognition of the user's position and gesture (finger pointing), aims at supporting people with visual impairment in orientation, navigation, and mobility tasks in known- or unknown-areas.

The main aim of the SoundingARM system is to avoid, where possible, the locomotor/tactile exploration of an environment by providing essential spatial information: often the blind user only needs to know the type of environment or whether a room contains a specific object that he/she needs or wants. This system may be used, for example, in order a) to provide information about the type of products placed in the indicated shelves in a shop, b) to offer blind clients the "first sight" of a hotel room, or c) to facilitate the reintegration into private living spaces of a subject who acquired visual impairment in adolescence/adulthood [15].

With SoundingARM, users need only to carry out two tasks: enter a room and, standing in the doorway, point in a direction in space with their arm. The system replies supplying auditory information about the objects placed in the indicated direction. These simple tasks help blind people develop an auditory map that can be used:

- to immediately recognize the type of environment;
- to quickly detect a specific object;

- to safely move in space.

SoundingARM, on the user point of view, might be used with the following objectives [15]:

- facilitating the reintegration into private living spaces of a subject who acquired visual impairment in adolescence/adulthood;
- encouraging the conversion of the "visual/spatial representation of space" into "an area-of-action representation of space";
- fostering the evolution from the condition of total non-self-sufficiency through the acquisition of proficiency and safety in mobility and orientation tasks;
- facilitating the decoding of environmental information and their organization in a spatial reference system;
- promoting the gathering of multi-sensory information about the environment.

5.2.3 The System Architecture

In general, SoundingARM employs the Microsoft Kinect sensor to determine the user's location as well as a software that utilizes the localization data to generate auditory information about the space. SoundingARM does not require sensors to be worn, pointing devices to handle, or tags to mark the environment. It uses instead standard hardware such as the Microsoft Kinect sensor, a normal Windows 7 computer, and a software to manage the localization data in order to generate vocal information about the space.
In particular, the overall system architecture (Fig 5.4) of SoundingARM is constituted by:

- *Input components* (in blue), a) a Microsoft Kinect sensor, connected by USB up to a computer with Windows 7 and b) a configuration file, containing information about the furniture of the room;
- *Real-time processing components* (in red), which are: a) the Microsoft Kinect SDK (Software Development Kit) for the recognition of the skeleton, b) the Server Application that analyzes the user's skeleton and understands what object the user is indicating, and c) the Pure Data patch which manages the audio rendering;
- *Audio output components* (in green), loudspeakers and headphones.

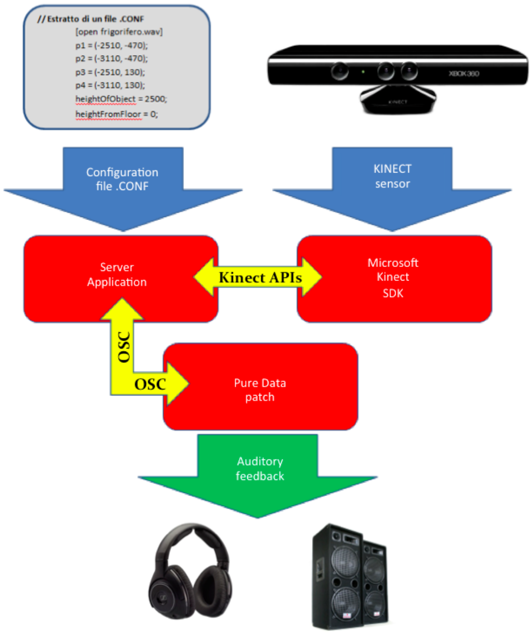

Figure 5.4: SoundingARM - The system architecture of SoundingARM.

5.2.4 Usability Testing

The SoundingARM system, on March 6, 2012, was tested by this author with ten adults with different degrees of visual impairment. Some of them had also motor/cognitive disabilities associated. The preliminary usability test was informal/qualitative and carried out in a kitchen of the Regional Institute "Rittmeyer" for the Blind (in Trieste, Italy). This environment was unknown to the majority of them. The overall purpose of the preliminary testing was to verify if the SoundingARM system performs at an acceptable level for these representative users. It is necessary, indeed, that the system firstly be tested by users in order to ensure that they have a positive and efficient experience in using it.

Identified testing participants received instructions prior to the beginning of preliminary testing in order to have them familiarize with the tasks which were: a) enter the room and, standing in the doorway, b) point the arm in several directions in order to identify the static objects.

They had to identify twelve static objects; a table, a fridge, a wall cupboard, a dish-rack, an extractor fan, a cooker, a sink, a switch, two radiators, and two windows. Test-takers did not know how many objects they had to localize [15]. In this case, the system used a speech feedback, in particular, a Text-to-Speech synthesis of the object description, or simply, its name.

5.2.5 Discussion on the Observations

The following remarks were collected by means of the test-takers' direct observation in order to obtain information about a) their approach to the system (their actions, potential difficulties, comments) and b) the system performance [15].

- Blind users with cognitive impairments associated requested: 1) further verbal explanations compared to those scheduled; 2) a "physical guide", given that the pointing gesture has to be as precise as possible; because of the short duration of the familiarization phase, these users were allowed to ask for help in order to perform the proper movement, in fact, they were guided by an external operator whose body, however, interfered with the user's skeleton recognition, leading SoundingARM to make errors; 3) a tactile signal on the floor in order to maintain the position on the doorway (they tend to enter the room as they do usually).

- Blind test-takers using the wheelchair did not localize all the static objects; this occurred probably because wheelchairs constituted an obstacle for the skeleton recognition and the head reference point was almost on the same level of the Microsoft Kinect sensor.

- Blind test-takers with no other associated impairment had easily identified all the objects of the room.

Generally, blind people never localize any target (acoustic or not) by pointing to it, an action more related to vision than to hearing/touch [90]. However, finger pointing is a basic communication task in interpersonal communication for blind persons too, indeed it is taught in orientation and mobility courses. The SoundingARM system, in this case, may represent a useful tool to improve this ability.
Moreover, when trying to reach a sonorous object, a blind person usually gropes for it rather than reaching for it directly [90]. In fact, the ability to reach a specific point directly is very difficult to acquire without visual reference points. The development of this important ability is obtained by means of the construction of a cognitive map, a spatial mental model [80; 81] which, in this context, is fostered by the acoustic feedback given by the SoundingARM system.
Overall feedback regarding SoundingARM was positive (informal interview). All participants indicated that they would enjoy using the system and that their performance would improve with more time to practice using it.

5.2.6 Preliminary Study

The SoundingARM system, in October 2012, underwent two further evaluations performed at the Italian Union of the Blind and Partially Sighted (UICI) of Rovigo, Italy.

The participants, 11 males and 9 females aged between 18 and 71, were selected in order to have 4 groups of 5 persons each, uniform in terms of visual-motor skills.

The two preliminary tests, summarized in Fig. 5.5, have been administered by Daniele Marabese – who also developed the last version of the software – during his master's thesis work supervised by Antonio Rodà, Federico Avanzini, and Serena Zanolla [91].

Test 1: SoundingARM vs tactile-locomotor	
Task	Measurements
To explore the room	Time of exploration Number of identified objects
To reach three objects	Time taken to reach the objects

Test 2: Auditory-icons vs Speech icons	
Task	Measurements
1 minute to explore the room	
To point and reach three objects	Time taken to reach the objects

Figure 5.5: The preliminary study - Two further evaluations of the SoundingARM system have been administered in October 2012.

5.2.6.1 The First Test: Consistency of the Spatial Map

The first test was administered in order to compare two different strategies of spatial map creation: using the SoundingARM system and by locomotor-tactile exploration. The participants performing the first test consisted in 10 visually impaired users of whom 5 formed the experimental group (using the SoundingARM system) and 5 the control group (locomotor-tactile exploration).

Participants were presented with the following main task: to explore, regardless of a time limit, an unknown kitchen environment without receiving any information regarding the location and nature of the objects to be individuated within the space.

The auditory feedback supplied by the SoundingARM system consisted in Speech Icons (a real voice

recording of the names of the objects). All the sound cues used in this test, and the one that followed, have been stored as *.wav* audio files and, as they do not depend on the SoundingARM software, they can easily be switched and modified.
The first test consisted in two tasks to be performed in two successive phases:

1. individuate the position of as many objects present as possible; the experimental groups performed the exploration starting at the threshold of the room and using SoundingARM, while the control group conducted the locomotor-tactile exploration by entering the room;

2. reach 3 target objects present in the room: the roll of kitchen paper, the coffee maker and the chair; the experimental group performed this task without using SoundingARM.

For the first task, it has been recorded (for each test-taker) a) the number and type of the identified objects and b) the total length of time it took to complete the exploration.
The eight objects to individuate were: the coffee maker, the roll of kitchen paper, the cooker, the oven, plates and tableware, the rubbish bin, the chair and the cutlery.
For the second task it has been recorded a) the total length of time it took each user to reach each of the three target objects, and b) the total length of time of the task (sum of the three spans of time).
The aims of the test were to assess:

1. which of the two groups, by average recognizes and correctly locates the highest number of objects;

2. which of the two exploration methods is more effective to build a spatial map.

5.2.6.2 Analysis of the Results

In order to establish if the two treatment groups were statistically different, the Mann-Whitney U test was used, since it is a non-parametric statistical test for assessing whether two samples of independent observations tend to have larger values than the other; the comparison is made using ordinal data and the population, from which the samples are drawn, have similar distributions [13]. Moreover, this test is used for very small samples (between 5 and 20).
As shown in figure 5.6, both groups performed good results in the exploration of the kitchen environment but no statistically significant difference can be found between the two groups ($p > .05$).

Task 1: the experimental group reached an average of 7.4 object out of a total of 8, and this is an excellent result considering the fact that objects such as the roll of kitchen paper and the coffee

Results test 1

		task 1		task 2			
genre	age	n objects	time [s]	paper towels [s]	coffee [s]	chair [s]	tot time [s]
M	47	8	144	83	14	20	117
F	27	7	140	81	10	9	100
F	59	8	53	N/D	18	16	N/D
M	52	6	110	5	30	33	68
F	23	8	90	5	81	27	113
F	19	7	58	24	32	9	65
F	47	8	121	16	7	65	88
M	65	8	196	N/D	15	29	N/D
M	71	8	192	13	17	6	36
M	24	8	142	4	18	6	28

soundingARM

tactile-locomotor

	soundingARM	tactile-locomotor
Average number of identified objects:	7.4	7.8
Average time to explore the room:	107s	142s
Average time to reach the objects:	99s	54s

No statistically significant difference can be found between the two groups (p>0.05)

Figure 5.6: Consistency of the spatial map - The results of the first test showed that both groups performed good results in the exploration of the kitchen environment.

maker are relatively small; the control group managed to reach a total average of 7.8 objects out of 8, a higher performance compared to that recorded for the experimental group.

One of the evident advantages of the SoundingARM system is a substantial decrease of the length of time necessary to complete the exploration of the environment. It was recorded an average time of 107.4 seconds for the experimental group, about 30% lower than the average time required by the control group.

Task 2: analyzing the time lengths required by all the test-takers to reach the indicated objects, it can be immediately noticed how, except in two cases (1 per group), all the users were able to reach the location of the required object: this shows that despite the exploration method, the users created a consistent spatial map.

The users of the control group, on average, were faster at reaching the first and second objects; while reaching the chair was performed more rapidly by the experimental group. As a whole, SoundingARM can be considered a good alternative to locomotor-tactile exploration as it allows the user to build a sufficiently accurate spatial map, in a short time, without having actually to move within the environment itself.

However, superiority of the tactile exploration over the SoundingARM system is evident, as the latter does not supply multisensory information. As for the improvement of the system, these considerations suggest that it would be necessary to integrate the feedback with further information on the objects present in the environment such as the distance and dimensions.

5.2.6.3 The Second Test: Auditory Icons versus Speech Icons

The second test was administered in order to compare the efficiency of the two different auditory feedback supplied by the SoundingARM system: Auditory Icons and Speech Icons. A further group of 10 visually impaired users participated in the test; the participants were divided into two groups of 5 users each: the experimental groups used SoundingARM with Auditory Icons, while the control group used SoundingARM with Speech Icons.
The Auditory Icons were made using real sound recordings of the target objects (e.g. oven, coffee maker, cooker); while for other objects, indirect sounds that identified the objects, for example the sound of the material the objects are made of (e.g. the table, the chairs, the plates, and tableware). In line with what Keller and Stevens [92] claim, it is often necessary to mix several sounds together that equally characterize the object in order to obtain a more direct and expressive icon (e.g. cutlery clinking against each other). The Speech Icons were the same ones used for the former test.
The test was conducted in two phases. During the first phase, each user was asked to explore the kitchen environment in 1 minute using SoundingARM. During the second phase, each user was asked to point to the direction of:

1. the cutlery drawer, and reach it;
2. the rubbish bin, and reach it;
3. plates and tableware, and reach them.

The aim of this test consisted in assessing how efficient the Auditory Icons are, compared to the Speech Icons, in aiding the user to recognize/learn the position of the objects within space. Through this test it were recorded a) individuation or not of the position of the objects within a space, and b) the time required by the user to reach the objects.

5.2.6.4 Analysis of the Results

As shown in figure 5.7 the results of the second test attests that both sound cues appear to be efficient: no statistically significant difference can be found between the two groups ($p > .05$, Mann-Whitney U test). In some cases, the users were not able to precisely indicate the correct direction and nevertheless reach the object successfully. As for the average lengths of time required to reach the objects, the participants that used the Speech Icons appeared to be faster, though the average advantage is merely by 20%.

Results test 2

genre	age	cutlery [s]	trashcan [s]	plates [s]
F	37	8	N/D	10
F	44	12	18	4
M	69	N/D	14	9
M	54	9	7	3
M	20	10	9	2
F	21	1	2	1
M	28	14	7	1
F	66	8	11	3
M	42	12	23	4
M	18	7	9	2

auditory icons

speech icons

Average time to reach the objects: 27s 21s

No statistically significant difference can be found between the two groups (p>0.05)

Figure 5.7: Auditory Icons vs Speech Icons - The second test has shown that both sound cues appear to be efficient.

5.3 Memory Sonoro: an Audio-Based Augmented Reality System

5.3.1 Introduction

Memory Sonoro (Fig. 5.8) – conceived by this author and developed in collaboration with the Mathematics and Computer Science Department (DIMI) of Udine University [16] – is an interactive multimodal system that extends the techniques of Augmented Reality (AR) to the domain of directional audio. Memory Sonoro in fact uses optical tracking technology and audio spatialization techniques in real-time through speakers to develop an acoustic-tactile interface, providing a play/educational dimension, accessible to blind and partially sighted users.

Implementing these technologies allows users: a) to interface the system by means of real objects (cards with markers) and b) full freedom of movement (no need to wear sensors) within the three-dimensional sound space.

Figure 5.8: Memory Sonoro - A child who has been totally blind since birth playing a game of Memory Sonoro. Source: [16]

Memory Sonoro is an AR system that uses sound to provide the information, which would otherwise be inaccessible to the blind and partially sighted [93]. It is based on the very popular simple board game Memory, which consists in finding the pairs of identical cards. The pairs of picture of the classical Memory game, in this case are replaced by pairs of markers (placed on cards) that the system recognizes and links to couples of sounds. In Memory Sonoro therefore, the players must find the pairs of identical sounds. Moreover, the sounds can also be spatialized in real-time when the user shifts the tags on the horizontal or frontal plane.

Memory Sonoro belongs to the category defined as Audio-Based AR systems; in fact, its aim is to extend the realm of AR to sound. In Memory Sonoro the position of the objects within space is used to "augment" the sound scenario and modify the perception of the sound in real-time. Memory Sonoro is based on an optical tracking system that solely implements image analysis processes to obtain information regarding the objects shown on the board. The optical tracking is connected to markers (ARtags), bidimensional objects on which an image that can be recognized by specific image processing algorithms. This solution is easily implementable, as it does not need important investments in terms of hardware. The markers offer also the possibility to use a tactile medium.

5.3.2 Related Works

Learning plays an important role in auditory scene analysis and sound localization performance. A increasingly wider line of research focalizes its attention on audio-based interactive systems to enhance learning and cognition in children with visual impairments. In particular, these research investigations focus on sound-based systems designed to develop sound discrimination/localization skills. This approach is based on the assumption that, "while tactile devices like braille lines can only provide a limited amount of information per time, audio has the capability to provide a lot more information at once if made surrounding and spatial" [93]: the three-dimensional auditory interface increases the information between the real environment and the blind user.

The game industry does not have as its primary concern the adaptation of the user interface of new technologies to the ergonomics requirements of the visually impaired, whereas the interaction between the visually impaired children and these technologies is of crucial importance for their inclusion into the society of their peers [94].

Sound-based systems for visual impairments usually employ auditory information as the main output channel and haptic devices for input [95]. We can mention, for instance:

- AudioDoom [96], a sound-based computer application focused at stimulating and reinforcing spatial representation;
- AudioMath [97], an "audio-based interactive virtual environment" to enhance short-term memory and learn mathematics;

- Terraformers [98], an acoustic interface developed with the aim to bridge the gap in the game industry between the mass market 3D games for sighted and the sound games for blind and low-vision users;
- AudioChile and AudioVida [95], "3D sound interactive environments", which are used to enhance spatiality, immersion, and orientation.

With the aim at investigating the use of sound binaural perception for three-dimensional source localization, an audio-based version of the Memory Game, Memory Sonoro – accessible by the blind through a tactile user interface – has been devised [16]. The first version of the Memory Game for visually impaired was developed by Delić and Sedlar [94]; the goal of this system is to locate the pairs of objects/sounds presented on the computer screen. Objects are described by a speech synthesizer and selected by giving verbal commands or using the keyboard.
Figure 5.9 clearly shows that the horizontal position of the object is described by stereo presentations, while the vertical position is indicated by using different audio frequencies (pitch of the synthesized speech).

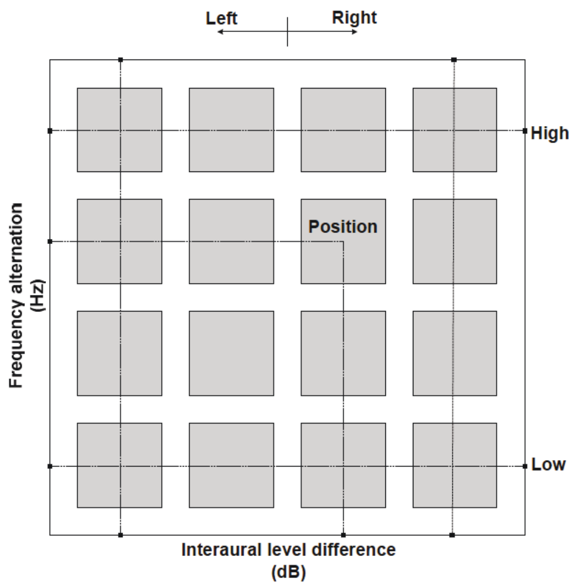

Figure 5.9: The Memory Game for bind people - The horizontal position of the object is described by stereo presentations, while the vertical position is indicated by using different audio frequencies. Source: [94]

Unlike the above version of the Memory Game, Memory Sonoro is a sound-based system which uses video-tracking techniques and tactile interfaces (cards). Memory Sonoro allows blind users to interface with the system by means of real objects (cards). The user plays with cards which, in this case, present pairs of sound instead of pairs of images, as in the traditional Memory Game [16]. By manipulating cards, the user discovers the objects (represented by sounds) and seeks for the pair of identical sounds. Moreover, the user receives a three-dimensional acoustic feedback according to the position of the cards in the game area.

5.3.3 Objectives of the System

The ability to orientate oneself and move within a space is the result of a specific sensorial education which allows the impaired to fully exploit the sensorial stimuli and use them to explore the environment, moving safely, understanding the position and the spatial proportion of the objects. Tactile-auditory sensations need to be exploited to help build mental images, perceive obstacles, orient oneself through an environment and enhance the ability to individuate reference points.

Memory Sonoro finds application exactly in this type of context positioning itself as a teaching aid for the blind, for the development and exploitation of the vicarious senses and the reduction of the secondary effects of blindness. In particular, Memory Sonoro can be used to:

1. develop tactile-auditory perception integration;
2. identify and analyze an object and its details;
3. learn exploration strategies and orientation within space;
4. denominate and describe sounds;
5. verbalize the logic processes and the operative and spatial strategies;
6. match by semantic connection (sound/object);
7. strengthen attention and memory abilities.

Memory Sonoro has a wide range of possible applications: it might be re-designed in order to trigger specific learning in children both with typical and atypical development. Moreover, it might improve or rehabilitate language, mathematical, and conceptual knowledge.

5.3.4 System Requirements

Memory Sonoro requires the following hardware components.

- A webcam (Fig. 5.8), positioned above the player/user, on a high-stand at about a meter from the plane of the game board so the direction of the framing is perpendicular to the floor: this setting a) leaves an ample liberty of movement to the user and b) allows a high level tracking process (avoiding distortions in the tracking of the markers position).

- The cards (Fig. 5.10) are the main element through which the users interact with the system. These consist in 8 by 8 cm cardboard wood bases. The surface where the paper marker is applied is smooth; while the back of the card is rough or velvety to the touch. This is necessary so that the user always knows which of the cards are face up and which are face down.

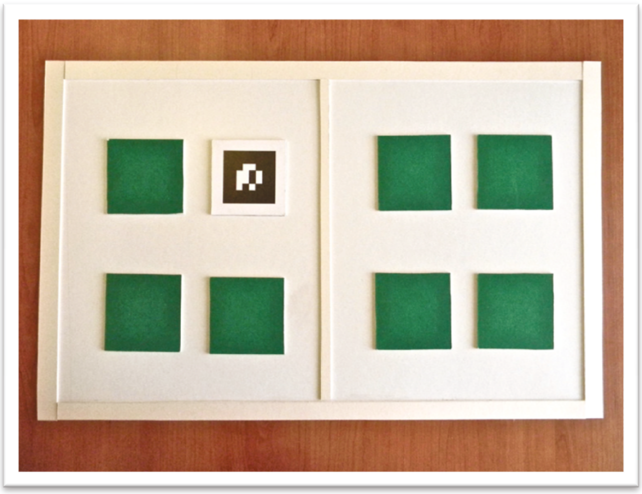

Figure 5.10: Memory Sonoro - The cards, the markers and the game board. Source: [16]

- The game area needs to be clearly delimited so that the players have the exact perception of the dimension of the game board. The game area consists of a 58 cm x 24 cm board that has the borders and the mid-line in relief (Fig. 5.10). This dimension guarantees that the markers always remain within the webcam and tracker field of view.

- A MacBook laptop computer, with the Mac OS X Operating system (in this case, Version 10.8.6).

- A soundcard.

- Six speakers in Dolby Surround 5.1.

5.3.5 The System Architecture

Figure 5.11 describes the architecture of the Memory Sonoro system.

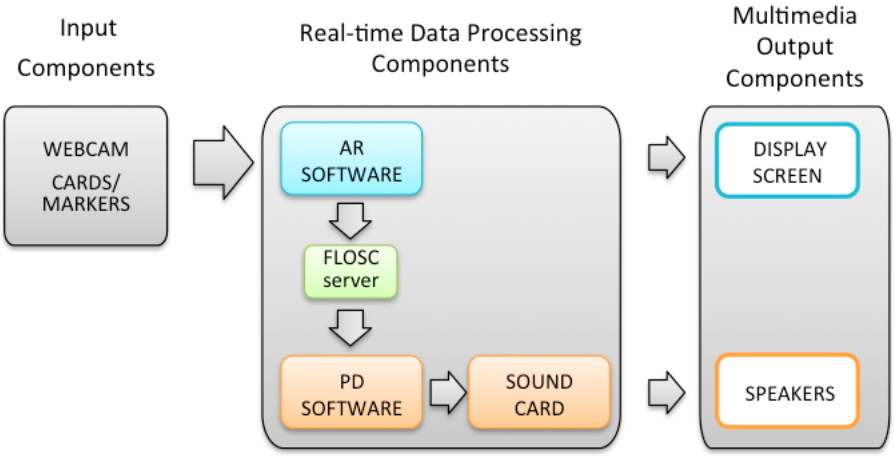

Figure 5.11: Memory Sonoro - The System Architecture of Memory Sonoro

Input Components. Memory Sonoro uses an optical tracking system: all the information regarding the position of the objects within a scenario derives from the analysis of video signal from the webcam. The user interface consists of pairs of cards (variable number of pairs). On one face of the cards there are the markers, black and white images that the system "recognizes" through an image-processing algorithm. The position of the cards on the game board determines the sound spatialization.

Real-Time Processing Components. An AR software (swAR), an application developed with Flash and the FLARManager library, that provides for: a) the optical tracking of the markers within the scenario, b) the overlapping of virtual images on the video taking of the real scenario and c) the conveying of the information on the markers to the audio software. Once the marker is localized, the swAR overlaps the virtual images (in 2D or in 3D): virtual colored circles, a large one in the centre and four smaller ones in the corners (Fig. 5.12). When two identical markers are set close together, a virtual colored line links the two centers.

An audio software developed in Pure Data (swPD) a) reproduces the sounds corresponding to the markers and b) spatializes them according to their position on the game board. When a new marker is introduced in the game scenario, the swPD executes the corresponding sound and if the marker is moved, the sound changes the source direction according to the new position.

The Audio Card manages the data flow sent by the swPD: the data are divided in 6 independent audio signals and sent to the speakers.

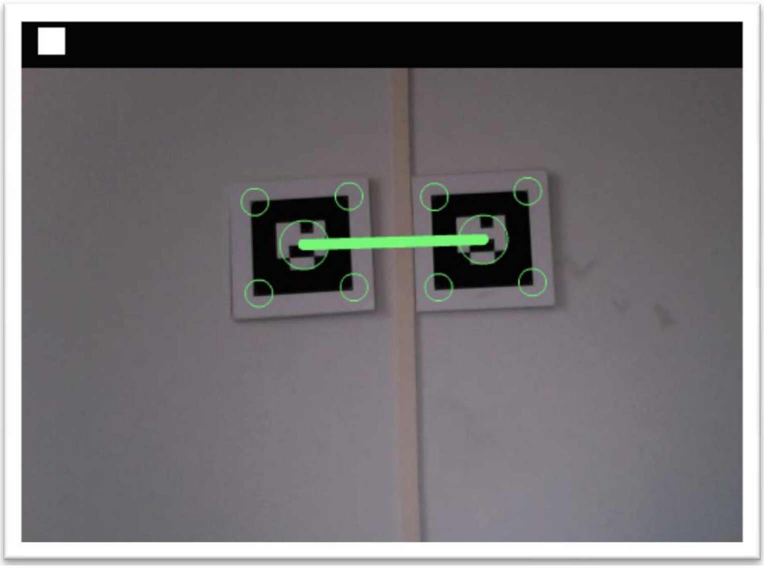

Figure 5.12: Memory Sonoro - Virtual graphical control elements. Source: [16]

Audio Output Components. The screen shows the real scenario of the game board and virtual images generated by the swAR overlap it. For a better control of the functioning of the system, it is possible to visualize the binary data that the digital version of the images comprises from which the swAR extracts the markers (Fig. 5.13). The six loudspeakers that create the spatialization of the sound need to be arranged in a 5.1 surround sound configuration: opposite, left, right, behind left and behind right; the subwoofer has to be positioned in the lower centre.

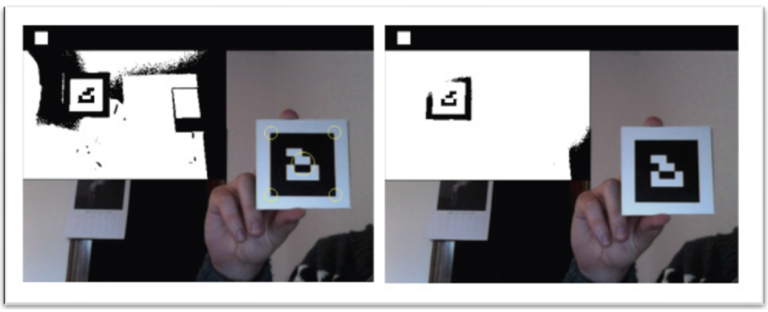

Figure 5.13: Memory Sonoro - Control Screen. Source: [16]

Memory Sonoro therefore, consists of an AR software component (swAR) that sends OSC messages to the software that performs an audio processing of the signal (swPD). These messages contain the information regarding the type of marker individuated and its position and are sent through a Flosc server (used precisely to send the OSC type messages to/from Flash). In particular, the AR component is managed by the FLARManager library for Adobe Flash. The FLARManager framework allows analyzing and tracking the markers captured by the webcam; the resulting data are then used to "augment the real scenario" with sounds and images. The virtual graphical elements are visualized on the screen and are generated directly by the Adobe Flash libraries, while the sound events are reproduced and spatialized by the swPD based on the information sent by the swAR through OSC messages.
In a nutshell, the main steps of this process are:

1. individuation of the marker present in the video flow;
2. comparison between the individuated markers;
3. overlapping virtual images on those captured by the webcam;
4. transmission of the OSC message to the server;
5. reproduction of the sound;
6. spatialization of the sound.

5.3.6 The Game

This paragraph describes how the users – the sighted helper that manages the system and the visually impaired child/adult who plays using the cards – interact with Memory Sonoro.
The main task of the helper is to verify in real-time if the system is working properly. For this reason, thanks to a display, the system shows a) the video flow of the game board with augmented graphics and b) the dialog box of the binarized images (Fig. 5.13) supplying the following information.

- The binarization threshold; the operator adjusts the threshold according to the illumination of the room by using the up/down arrow keys on the keyboard.
- The accurate detection of the markers; the window of the video flow with overlapping graphics allows to control the functioning of the system and whether the marker is correctly recognized: the fact that the graphics are visible indicates that the system is properly performing the recognition; when the graphics do not appear then there is a tracking problem that can be resolved

by a) modifying the threshold or b) resetting some elements of the installation, for example the illumination conditions or the position of the camera.

- The accurate detection of the pairs of markers; when the child/player draws the identical cards together (same sounds), the system links them virtually with an 8 pixel wide colored line that crosses the centre of the two individuated markers. The helper needs to verify that this type of graphics appears on the screen, and in case it does not he/she must modify the settings of the software or some other element of the installation.

The main user – the child with visual impairment – interacts with the system by manipulating the cards. Following, there are the actions performed by the child during the game:

1. turn over the card and listen to the corresponding sound;
2. turn over another card and listen to its corresponding sound; the action is repeated until the paired sound is individuated;
3. confirmation from the system; the child presents both cards to the camera by drawing them together. If the choice is correct, the system executes a sound to confirm the correctness. If the choice is not correct, the system does not emit any sound: according to the experts in education of the blind and visually impaired, an error sound is a redundant information that could invalidate the task of memorizing the principal sound stimuli.

5.3.7 Usability Testing

The prototype of Memory Sonoro has been tested on a sample of 12 partially sighted and blind people, aged between 4 and 45 years. A usability test with an assessment based on verbal feedback from the users, questionnaire, has been conducted [99]. Because the sample group of users was so heterogeneous (in terms of age, type of disability and ability to use vicarious senses), each assessment session was adapted according to every single subject (number of cards and timing).

The usability test offered the possibility to assess the system feedback, in order to evidence the possible faults in the development that could be resolved before implementing a large-scale research. Besides assessing the performance, the usability test allowed to have an understanding of what the users thought about Memory Sonoro; in fact, even though performance levels can be highly positive, the users opinion on the technology is equally important to help develop further features of the system. The usability test was useful to:

1. identify which actions were more congenial and which instead were more difficult to carry out;

2. observe limits and potentials of the system performing in conditions other than those imagined in the engineering phase.

For the test complex sound stimuli were used: the sound of a dive into the water, a helicopter, an airplane, a racing car, the sounds made by several animals and environment sounds. Following the guide lines of usability tests [100], we submitted two questionnaires: one to be completed before performing the test, which consisted in the description of the type of user (sex, age, education, visus and the level of familiarity with the Memory Game) the other, to be completed after the test, to assess the tactile and sound feedback given by the system and collect the various opinions and suggestions regarding the overall game experience.

5.3.8 Discussion on the Results

The usability test allowed to individuate the problems described below.

1. The users freely manipulate the cards, turning and touching them, hence the markers are covered by their hands obstructing the correct tracking; to compensate for this problem, the users always need to place the cards on the game board and take away their hands: non-spontaneous action that limit both the interaction with system and the spatialized-sound perception experience (in particular on the frontal axis). Placing the webcam under the game board (which in this case should be transparent), using for example the TUIO technology[1], could resolve this type of problem.

2. Once the players turn a card and listen to the corresponding sound, they usually leave the card face up and search for the pairing sound; moving around on the game board, often by passing over the first card again, re-activates the sound. This situation generates ambiguity and confusion to users. At the moment, this problem has been resolved by asking the user to turn the card face down immediately after listening to the sound. To assure a spontaneous use of the cards a system with a different functioning logic needs to be implemented.

3. Because of the difficult conditions of illumination (neon lights above the game board) the performance of the optical tracker have not been optimal. The solution to this problem, for the time being, was found by setting the binarization threshold and asking the users to move the cards to other areas of the game board.

4. All the users defined the game to be pleasant and easy to understand and play.

[1] http://www.tuio.org/

5. Only 4 users out of 12 declared that the spatialization of the sound actually helped them to find the sounds on the game board. All the other users, used spatial memory strategies to memorize the position of the cards and the corresponding sounds. Most probably the problem described in 1 also influenced the choice of which memorization strategy use: if the users had had the possibility to move the cards without limitations while listening to the sounds, they would have more clearly perceived the effects of the spatialization in real time.

Aspired future developments include, among others, the implementation of the modifications listed in paragraph 5.3.8. Moreover, as the technologies employed can perform functions that in this context have not been fully exploited – one of these being the possibility to send OSC messages in two directions – the purpose is to implement the communication of the data from the PD software to the AR software in order to offer a further element to control the functioning of the system. In order to facilitate the installation of the system, some modifications could be made on the control graphic interface, for example, add a graphic feedback relative to the connection status of the software components to the Flosc server.
The functions of Memory Sonoro can also be broadened by: a) using categories of complex sounds which can be associated according their meaning (e.g. the sound of water could be associated to the sound of a sink) or b) implementing a new variation of the game where the users task is to localize the sound source within a three-dimensional acoustic space at 360 degrees.

6

SoundRise and ParrotGameDiscrimination: IMEs for Production and Discrimination of Speech Sounds

6.1 SoundRise: Speech Recognition Technology in Education

6.1.1 Introduction

It is generally acknowledged [101; 102; 103] that interfaces with visual and multimodal feedback can facilitate the speech and vocalization education process for children with communication skills deficits by stimulating them to complete a task and by providing information about the acoustic properties of speech. These interfaces by employing speech technology are able to analyze human oral language as input and respond with different types of output (synthesized speech, images, and sounds).

With this in mind, SoundRise – a multimodal interactive application based on speech feature analysis – was conceived by this author and developed by the *Centro di Sonologia Computazionale* of the University of Padova, during the master's thesis work of Marco Randon and Stefano Giusto, supervised by Antonio Rodà, Federico Avanzini, and Serena Zanolla [104; 105].

The use of voice as an enactive instrument for learning is particularly interesting: the voice is the uni-

versal instrument for transmitting verbal and non-verbal information since it includes paralinguistic elements that convey emotions and moods ([17] and [18]). This becomes utterly important when the user, due to physical or cognitive impairments, is able to communicate exclusively through simple vocalic modulations.
SoundRise is an application that employs the voice to create visual signs which represent the main features of sound. The voice is graphically represented by the sun which goes up or down according to the pitch of the sound emitted, becomes bigger or smaller according to the energy of the vocal sound, changes color depending on the vowel uttered, and smiles all throughout the duration of the sound.

6.1.2 Related Works

Speech technology has recently become mature and available to the general public as well as capable of supporting advances in computer-based speech training (CBST) applications. At the same time, there has been an increase of the interest in the field of automatic speech recognition (ASR) in using speech technology to support learning for students, since ASR provides the means to interactively evaluate the speech of a learner on several educational dimensions. Therefore, the possibility to integrate sound, speech, text, videos, and animation has made possible to create systems able to significantly enhance the language awareness possible.
The research presented in this section is focused on systems that provide users with graphical feedback (visualization) according to their speech production. Nowadays indeed, the visualization of human speech in a graphic representation on a computer screen has become a feasible task [106].
The more recent graphics displays of speech visualization are capable to provide a meaningful feedback: it is possible to include, for instance, the animation of the mouth showing how the sound is to be produced [106]: Watch Me! Read (WM!R) [107] is a reading tutor with the aim to enhance literacy, comprehension, and communication. This software uses speech recognition techniques to assess student's performance and provide them with individualized visual and audio feedback (Fig. 6.1). In addition, the WM!R system calculates the number of words that the student has produced correctly. Moreover, teachers themselves can customize the contents of the system.
There are many other speech training systems, developed in the last decade, which follow this approach. Below, several systems using the graphical feedback to draw the speech in educational contexts are briefly described.
The SPECO system [108] is "an audio-visual pronunciation teaching and training tool for 5-10 year old children" with severe disabilities; this system provides an understandable and interesting visualization of the speech parameters in real time; the visual feedback helps children to see, therefore assess, whether

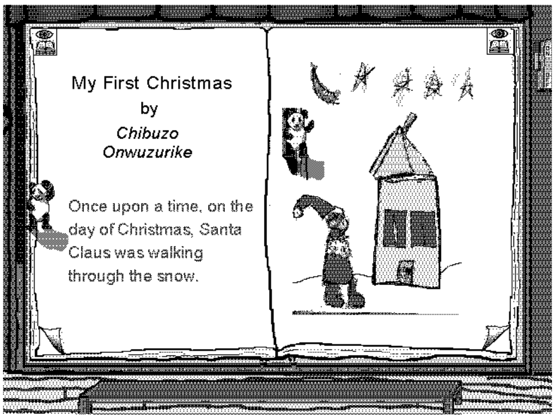

Figure 6.1: Watch Me! Read (WM!R) system - Reading view of Watch Me! Read showing a book written and illustrated by a schoolchildren. Source: [107]

their pronunciation is correct or not, and how far it is from the correct one (Fig. 6.2).

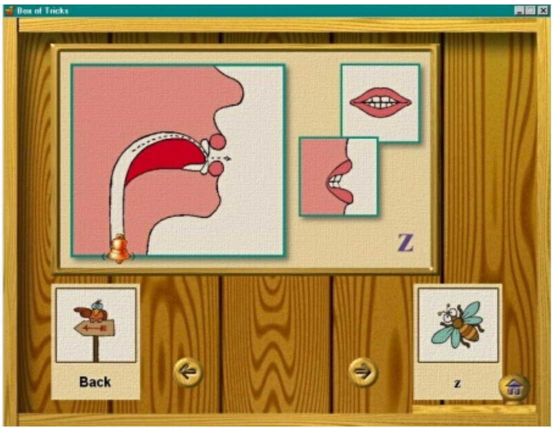

Figure 6.2: The SPECO system - Articulation picture for [z] sound; the ringing of the bell suggests that this vocal sound is "voiced". Source: http://www.enl.auth.gr/phonlab/box_of_tricks.html

The VoiceDraw [109] is a "voice-driven drawing application"; it uses the voice as a modality for creating drawings without manual interaction; the aim of this system is to enhance the creative expression of people with motor impairments which cannot use a mouse or perform other control tasks (Fig. 6.3). It is interesting to notice that this system utilizes non-speech vocalizations, that is to say, vocal sounds that do not correspond to words or phrases; these vocal sounds are non-speech properties of the voice: pitch, volume, and vowel quality that users can manage. Human voice is characterized by several features which people are able to control changing the vocal emission: the prominent are intensity, spectral centroid, a voiced/unvoiced flag, and the pitch.

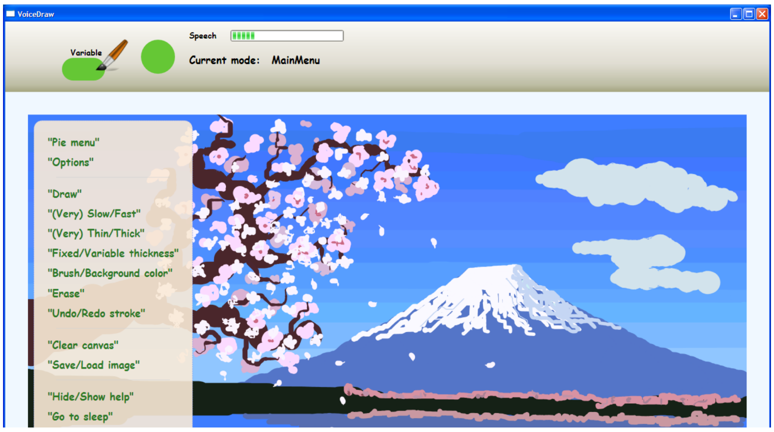

Figure 6.3: A screenshot of the VoiceDraw application. - The user with motor impairments creates a painting using his/her voice. Source: [109]

The speech recognition technologies can be used in very different educational contexts. The WinSingad system [110], for instance, has been used to monitor the vocal performance of a singer, children, or a choir. The system implements a set of displays (Fig. 6.4) which are organized in individual panels designed to provide a simple visual feedback (e.g. spectrograms, vocal tract area, fundamental frequency contour, input waveform, and spectral ratio).

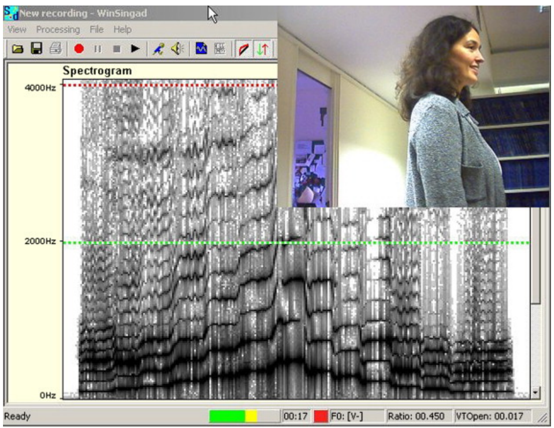

Figure 6.4: The WinSingad system - WinSingad spectrogram display for an arpeggio on [a:]. Source: http://www.sciencedirect.com/science/article/pii/S0892199705001323

The Voice Painter [111] stimulates the user to use the voice features and movement to draw visual signs on a canvas (screen). The focus of this system is to exploit the most relevant features of speech and map them into graphic features. This system has the potential to be used as an assistive technology in the field of speech therapy in order to reinforce vocalization and speech skills (Fig. 6.5).

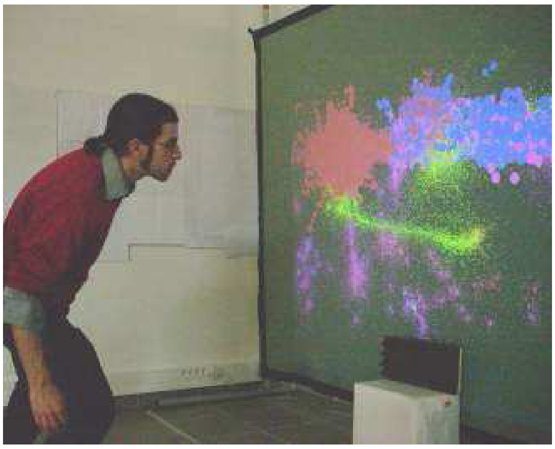

Figure 6.5: The Voice Painter System - Drawing by means of the Voice Painter. Source: [111]

On the basis of the systems briefly described above, SoundRise, the research project presented in this section (6.1), is focused on designing, developing, and validating IMEs using speech recognition technologies in real contexts; this IME is able to provide graphical feedback of every single feature of the voice: pitch, intensity, timbre, and duration which are very abstract concepts for pupils.

6.2 The SoundRise Application

On the basis of the existing approaches (Sect. 6.1.2), it is well-known that computers and Human Computer Interaction (HCI) techniques can be used as a method to help children by teaching and reinforcing both vocalization and speech skills. In this section, it is explained how these new techniques may be particularly exploited in order to enhance primary music education which needs to employ more specific and innovative tools.

Generally, music education is not given as much attention as it should, since it is often considered as an additional subject included in the curriculum and not really as a main subject. It is the exact opposite. Music education at primary level is very important also considering the role that music plays in today's society ([112]).

These are the main reasons why this author conceived SoundRise (Fig. 6.6), an application developed by the *Centro di Sonologia Computazionale* of the University of Padova during the master's thesis work of Marco Randon and Stefano Giusto, supervised by Antonio Rodà, Federico Avanzini, and Serena Zanolla [104; 105].

Voice is characterized by several parameters which human beings are able to control; the prominent features taught at a primary school level are; intensity, pitch, timbre, and duration. From the beginning

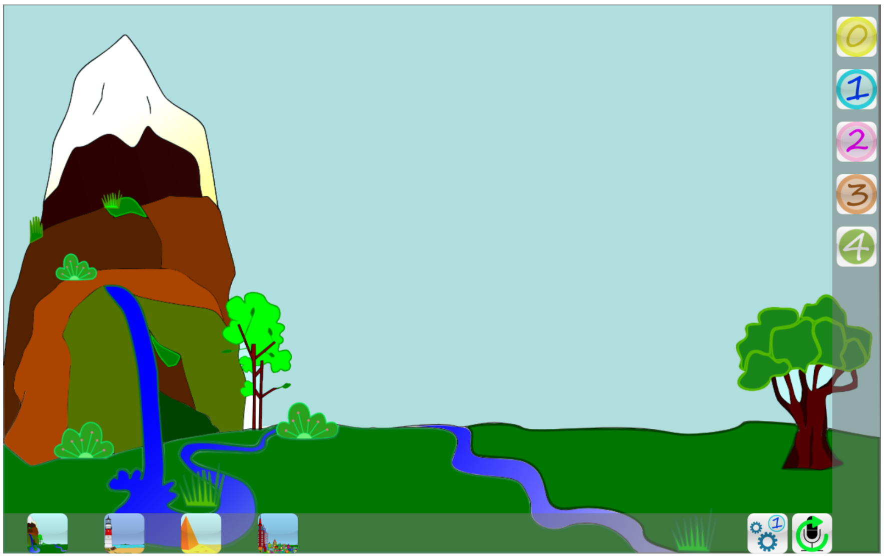

Figure 6.6: The SoundRise interface - SoundRise application creates an amusing graphical landscape.

of the research, the main idea was to create an interactive and multimodal system that was capable of mapping these vocal features into well-recognizable graphic signs, such as size, colour and position in real time. The goal was to teach pupils the vocal parameters and how to control them.

The outcome of this research is SoundRise, an interactive and multimodal application based on speech recognition technologies, able to provide graphical feedback according to each feature of the vocal signal. In particular, the SoundRise application is capable of using speech signals, captured by a microphone, to create consistent visual signs on a computer screen.

As well-illustrated in figure 6.6, the SoundRise application creates an amusing graphical landscape on a screen; by clicking the graphical icons at the lower left side of the screen, pupils can change landscape. The icon containing the microphone, at the lower right, is useful to calculate the noise environmental threshold, a very important function to discriminate silence from the users vocal sounds and to obtain the most efficient performance of the system possible, given that a school environment is usually rather noisy.

The *setup* icon at the lower right of the screen is useful to capture a users reference tone on which the application bases the extraction of the pitch. Selecting the setup icon the user is asked to pronounce the vowel /a/ for two seconds. The application stores the value of the pitch in order to correctly visualize the graphical information and to perform the voice analysis most accurately as possible.

A yellow sun appears above the horizon when users make a vocal sound. By selecting the buttons on the right, users can choose which voice parameter visualize (fig. 6.7). In detail, selecting the button

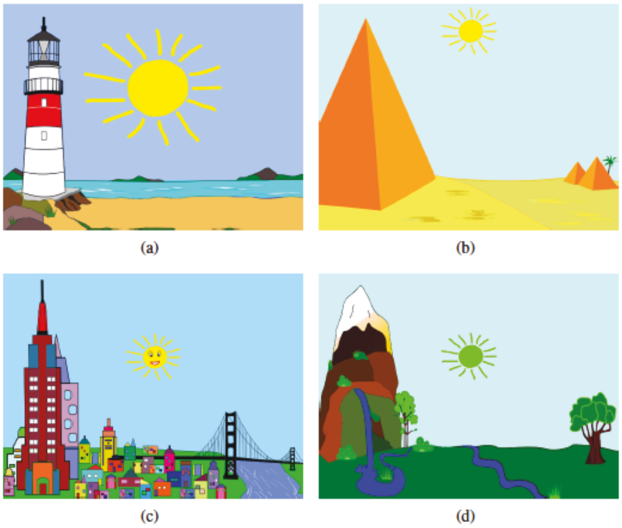

Figure 6.7: Voice features - The four visualizations of the sun provided by SoundRise according to the voice features extracted: (a) intensity, (b) pitch, (c) duration, and (d) timbre. Source: [105]

number:

- 0, the system analyses the amplitude of sound; this feature is graphically represented by the variation in the size of sun which becomes bigger or smaller according to the magnitude of the sound pressure; the user can enlarge/reduce the size of the sun making vocal sounds of different amplitude;
- 1, the system analyses the frequency of sound; this feature is graphically visualized by the variation in the position of the sun which higher or lower above the horizon according to the number of air pressure oscillations per second; the user, in this case, can move the sun up or down making vocal sounds of different pitch;
- 2, the system analyses the duration of sound; this feature is graphically represented by a broad smile of the sun;
- 3, the system analyses the spectral content of the vocal sound; this feature is graphically visualized by a colour variation; the user changes the colour of the sun by making different vowels;
- 4, the system analyses amplitude, frequency, duration, and timbre and visualizes them together; by means of this function the user can draw all the voice features together.

The SoundRise application implements a simple and user-friendly interface in order to avoid distraction on the part of the young user. Each landscape has a reference element – a lighthouse, a pyramid, a mountain, and a skyscraper – on the left to facilitate the evaluation of the graphical variations on the part of the user. The mapping between the auditory feature of vowel timbre and the visual feature of colour (function 3) is based on [113], a study on letters-colors associations in persons both with and without grapheme-color synesthesia: even if, generally, the association letters-colors is partially affected by environmental biases, the identified mapping vowel-color is the following: red for /a/, green for /e/, blue for /i/, orange for /o/, and grey for /u/.

6.2.1 The System Architecture

The SoundRise overall system architecture is showed in fig. 6.8. The system is constituted by the following main components.

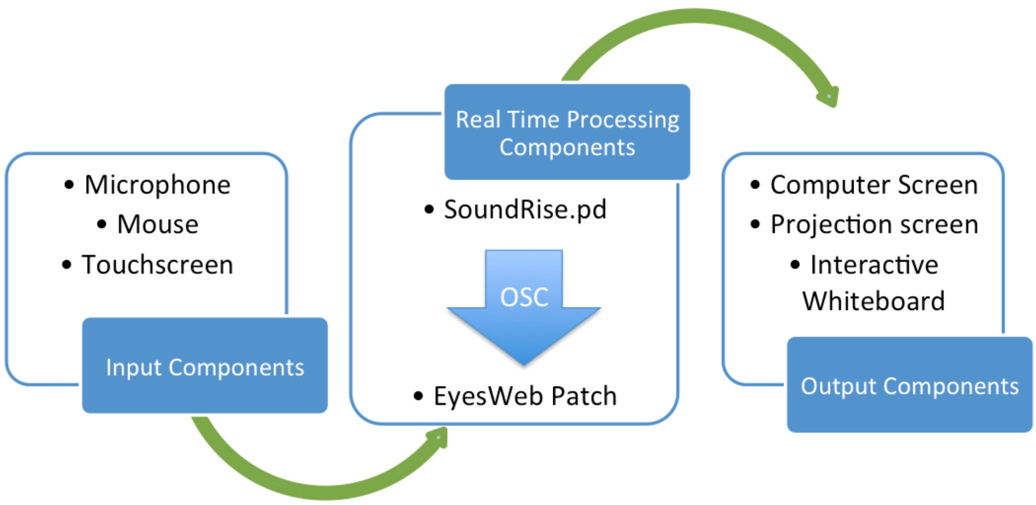

Figure 6.8: SoundRise - The overall system architecture.

Input components: a microphone (external or computer integrated) to capture the voice signal, a mouse or a touchscreen to select the various functions;

Real time processing of vocal signal components: a software patch developed in Pure Data (*SoundRise.pd*) enabling the feature extraction and the mapping (voice features-graphical features);

Multimodal Component: a software patch developed in EyesWeb platform[1], used in order to extend the level of multi-modality of the system;

Output Components: a screen for the graphical rendering (the computer screen, a projection screen, an interactive whiteboard).

The *SoundRise.pd* patch included several sub patches: a) the *pd settings* sub patch that manages the graphical visualization by means of the GEM (Graphical environment for Multimedia) library, b) *SoundRiseCore*, an abstraction of the *SoundRise.pd* patch, which contains the feature extraction algorithms, and c) other sub patches for the management of voice features, among them, the *pd vowel* patch which enables the vowel sound identification.
The SoundRise system has been designed considering the possibility to extend the level of multi-modality by means of the integration with the Stanza Logo-Motoria system. The combined usage of the two systems augments the interactivity of pupils' vocal experience enabling them to explore vocal features by moving within space. The user can select which feature the system has to extract by entering specific areas of the Stanza Logo-Motoria. In this case, the Stanza Logo-Motoria space is subdivided into 5 areas as showed in fig. 6.9. The protocol OSC (Open Sound Control) enables the communication between SoundRise and the Stanza Logo-Motoria system.

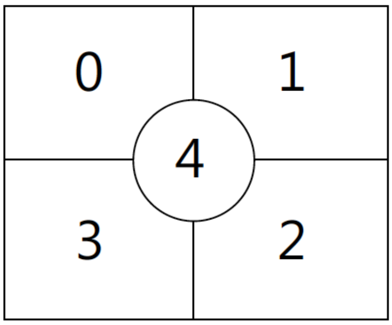

Figure 6.9: SoundRise and the Stanza Logo-Motoria - The Stanza Logo-Motoria in SoundRise modality. Source: [105]

If the user occupies areas 0, 1, 2, 3, the system triggers the vocal feature extraction and the graphical visualization of, respectively, intensity, pitch, duration, and timbre; even if the user occupies the central area (4) the system provides the vocal feature extraction and the relative graphical visualization of all the voice parameter together.

[1] http://www.infomus.org/

By means of the *libpd* library, the SoundRise prototype, initially developed in Pure Data, has been converted into a multi-platform application. The SoundRise application has been tested on several operating systems such as, *Apple Mac OS X, Apple iOS, Microsoft Windows XP, and Microsoft Windows 7*. Moreover, the use of *openFrameworks*, a toolkit for the development of multi-platform and multimedia applications, guarantees compatibility with other operating system such as, *Google Android* and *GNU/Linux*. This final version of SoundRise has a modular structure which allows the experimentation of different audio analysis technologies focused on rehabilitation of speech impairments.

6.2.2 Objectives of the System

The aim of the SoundRise system is to teach primary schoolchildren the main features of voice: pitch, intensity, timbre, and duration. Specifically, the intention is to use contingent visual feedback to a) motivate and reward vocalization and b) provide information on the acoustic properties of vocalizations. The challenge for teachers is to motivate pupils in order to engage them in tasks which interest and spark their curiosity.

SoundRise is an innovative method to teach pupils the features of voice by means of the exploration of their vocal skills. Autonomously, pupils discover, interpret, and understand the information given by the graphical feedback. Pupils have an active role since they control, by means of their voice, the graphical evolution of the feedback.

6.2.3 Usability Testing

In order to test the usability of the SoundRise application, two usability tests have been administered: one focused on measuring the system performance and one aimed at verifying whether the interface is intuitive for the user.

As regards the system-performance test, an heterogeneous group of test-takers was formed: a total number of 27 participants, both male and female, aged between 4 and 68 years. Test-takers were asked to perform several simple tasks; the administrator had to observe and write down whether the system provided the consistent feedback. The tasks, expressed orally by the test-administrator, are listed below.

1. "By means of SoundRise you can draw a sun on the screen. Try it". Here the test-administrator has to select the function 4.

2. "The icons at the lower left allow you to choose another landscape on which you can draw the sun. Try it." Once again, the function 4 has to be used.

3. “Select the setup button and follow the instructions (make the vowel /a/ until the button changes colour)”.
4. “Select the 0 button, produce a loud sound, and observe what happens to sun”.
5. “Now, make a soft sound, and observe what happens to the sun”.
6. “Select the 1 button, make a high-pitched sound with your voice, and observe what happens to the sun.”
7. “Now, make a low-pitched sound with your voice, and observe what happens to the sun”.
8. “Select the 2 button, make a long sound, and observe what happens to the sun”.
9. “Now, make a short sound with your voice, and observe what happens to the sun”.
10. “Select the 3 button, make a long vowel with your voice, and observe what happens to the sun”. This task is repeated for each vowel.

The results of the performance test are collected in fig 6.10. The answer “Yes” means that the system performed correctly; the answer “No” means that the system performed incorrectly. It is evident that the system failed in particular on the vowel identification function. This is probably due to the heterogeneity of the treatment group: the voice features database used by the system was realized with voices of children aged 9 and 10. The system almost never recognized the vowel /u/ and performed better only with female voices whose fundamental frequency fluctuates within ranges similar to those of children’s voices. These results show that it is necessary to widen the database of the vocal features and improve the vowel identification method.
As regards the usability test aimed at verifying whether the interface is intuitive for the user, a treatment group was formed: a total number of 39 pupils aged 8. As for the previous test, each test-taker was asked to perform some simple tasks; furthermore, they had to answer questions focused on verifying the comprehension of the graphical features (the graphical variation of the sun) and the awareness of the use of their voice. The tasks to perform were the same of the test described above. The questions to answer were the following:

- What did the sun do?
- What did you do with your voice?

N. test	Age	G.	Intensity	Pitch	Duration	A	O	E	I	U
1	27	M	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
2	28	F	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
3	28	M	Yes	Yes	Yes	No	Yes	Yes	Yes	No
4	30	M	Yes	Yes	Yes	No	No	No	No	No
5	4	M	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
6	31	F	Yes	Yes	Yes	No	Yes	Yes	Yes	No
7	29	M	Yes	Yes	Yes	No	No	No	Yes	No
8	27	M	Yes	Yes	Yes	No	No	No	Yes	Yes
9	30	M	Yes	Yes	Yes	No	Yes	Yes	Yes	No
10	27	M	Yes	Yes	Yes	No	Yes	Yes	Yes	No
11	23	M	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
12	23	M	Yes	Yes	Yes	No	No	Yes	Yes	No
13	30	M	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
14	27	M	Yes	Yes	Yes	No	Yes	Yes	Yes	No
15	19	M	Yes	Yes	Yes	No	Yes	No	Yes	No
16	24	M	Yes	Yes	Yes	No	Yes	Yes	Yes	No
17	57	F	Yes	Yes	Yes	No	Yes	Yes	Yes	No
18	31	F	Yes	Yes	Yes	No	Yes	Yes	Yes	No
19	25	F	Yes	Yes	Yes	No	Yes	Yes	Yes	No
20	28	M	Yes	Yes	Yes	No	Yes	Yes	Yes	No
21	14	F	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
22	50	M	Yes	Yes	Yes	No	Yes	Yes	Yes	No
23	48	F	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
24	17	F	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
25	68	M	Yes	Yes	Yes	No	Yes	Yes	Yes	No
26	28	F	Yes	Yes	Yes	No	Yes	Yes	Yes	No
27	63	F	Yes	Yes	Yes	Yes	No	Yes	No	No

Figure 6.10: Usability Testing - The results of the system performance test.

This test assessed whether the graphical feedback is consistent with the content to transmit; to better explain, when the educational objective is to acquire awareness by modifying the pitch of a vocal sound, it is necessary that the graphical feedback is fully consistent with the vocal parameter (in this case, the pitch). The results of the test showed that, for each single feature (functions 0, 1, 2, and 3), the graphical information is immediately comprehensible (fig. 6.11).

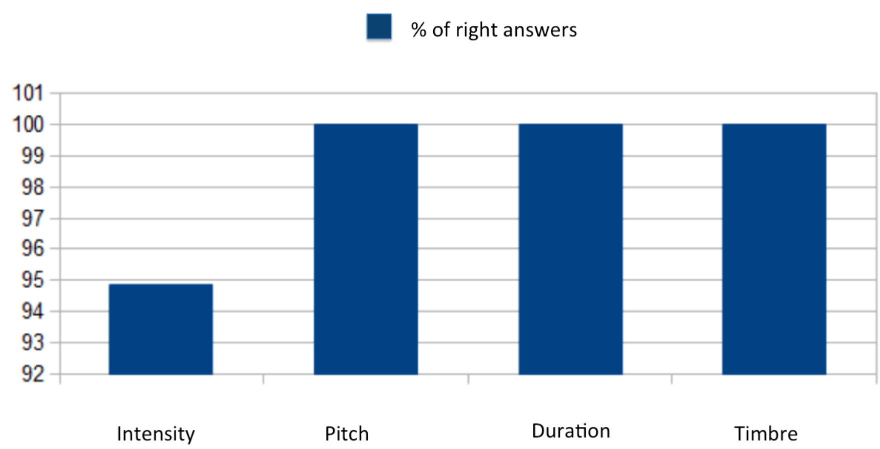

Figure 6.11: Graphical feedback comprehension test - For each single feature the graphical information is immediately comprehensible.

For the 4 function, where the sun shows all the features together, not all the graphical visualizations were identified. In the most cases, users identify 2/4 graphical feedback, this is probably due to information overload (fig. 6.12).

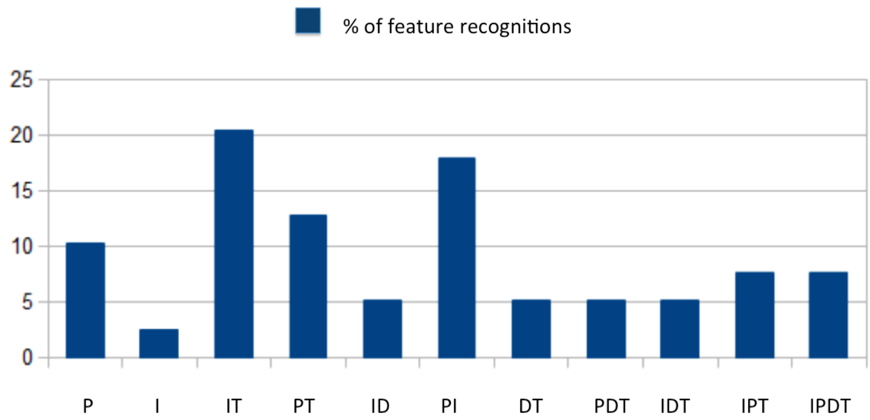

Figure 6.12: Feature recognition test - In the fourth function, the majority of users identify only 2/4 graphical feedbacks. P = Pitch, I = Intensity, T = Timbre, D = Duration.

Regarding the awareness of the control of the graphical feedback by means of the voice, it is possible to observe that timbre is associated with the less intuitive graphical feature (fig. 6.13). To be more clear, it is not immediately comprehensible that the colour of the sun is modified according to the vowels emitted. This is due to a) the vowel identification method implemented which did not provide a coherent colour feedback according to each single vowel and, once again, b) to the database of reference vocal sounds.

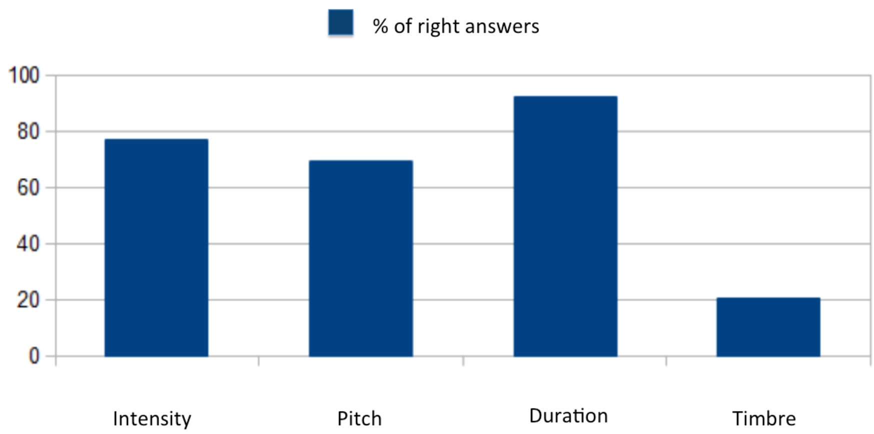

Figure 6.13: Timbre awareness test - Pupils did not understand that the sun changes colour according to the different vowel sounds emitted.

6.3 ParrotGameDiscrimination: a Multimodal Interactive Video-Game for the Evaluation of Speech Discrimination

ParrotGameDiscrimination (PDG) is a multimodal video-game for mobile devices – tablet, in particular – that was designed by the CSC of the University of Padova in collaboration with the DPSS of the University of Padova and this author in order to enhance the evaluation techniques of speech discrimination. In particular, the software was developed by Alessandro Dal Corso and Giacomo Baruzzo during the *Informatica Musicale* (Music Informatics) course within the master degree in Information Engineering at the University of Padova.

The PGD is an application that allows capturing the real capabilities of children to discriminate different sounds and speech fragments.

The ability to correctly and accurately interpret oral messages from others is essential for interpreting social interactions. Besides the difficulties with the possible social interactions, children with auditory processing concerns are likely to have severe difficulties in the class environment. The PGD application is designed to deal with these problems: the early assessment of children's capabilities to discriminate, for each sound feature, the difference among a number of sound stimuli allows to prevent a social disadvantage.

Another important goal of this application is to deepen the scientific research about the auditory processing of sound in specific populations of children – such as children with Down's Syndrome (DS) – who, due to auditory processing disorders, encounter difficulties in speech production.

The following section (6.3.1) addresses the related works and the theoretical background on which the design and development of the ParrotGameDiscrimination – an innovative tool for the early assessment of speech discrimination skills – are based.

6.3.1 Theoretical Framework and Related Works

The verbal and language development process begins at an early stage of a child's life. Within the first 12 months of life, normally infants develop the ability to respond to voices around them and, by the age of 2 years, toddlers begin to verbally express themselves and follow simple prompts. When pre-schoolers present a speech delay or a communication disorder, they are likely to have difficulty in responding appropriately or avoid speaking altogether. For this reason, it is important to identify a speech or language disorder as early as possible.

Many cases of language delay, phonological impairment, Down's Syndrome (DS), etc., characterized by an impairment in speech production, might rely on difficulties in speech discrimination. Theories, which identify a causal relationship between the two, are solid [114; 115]. However, it is still difficult to fully recognize to which extent an articulatory difficulty, which characterizes for instance DS, is due to a difficulty in discriminating acoustic characteristics of language. For this reason, innovative fine assessment tools have to be developed.

Up to a developmental age of 18 months, hearing (for children with DS this period extends until they are aged 2 to 3 years) is usually assessed by distraction tests. "Distraction tests involve making a wide range of sounds, many of them normal everyday sounds and speech, out of the sight of the child and assessing the child's response. These sounds made cover a wide range of sound frequencies (pitch) and intensities (loudness)" [116].

From a developmental age of 18 months, speech discrimination tests are used alongside, for instance, reactometer tests. A child with DS achieves this stage at age of 2 and 3 years [116]. The ability to discriminate speech sounds of different frequencies is tested by the toy test [117]: the child is prompted to pick up various toys without looking at the tester's face. The test administrator has to produce different voice intensities and needs to constantly check the tone and strength of his/her voice to carry out a reliable test (Fig. 6.14).

Figure 6.14: McCormick Toy Test - The therapist prevents the child from lip-reading by covering their mouth with a card and then asks the child to identify various toys. Source: http://www.aviva.co.uk

By a developmental age of around 2.5 to 3 years, children are able to co-operate with performance tests which are used to assess the response to pure or modulated tones. "At the beginning of the test session, the child is taught to respond by some fun activity – such as putting a peg in a hole – whenever they hear a noise" [116] (conditioning). Children usually quite enjoy this test; on the contrary, children with

DS are likely to have difficulty in understanding the rules of the game or simply they do not co-operate. "At a variable but later developmental age, when the child can cope with headphones, pure-tones can be delivered - a procedure known as warble tone audiometry" [116]. Generally, the sound discrimination tests are administered to verify the perception of everyday speech sounds. The "ability to hear pure tones does not necessarily mean that a pupil is able to discriminate between the different sounds used in speech. For this reason, speech discrimination tests are always used alongside pure tone tests as soon as a child is old enough to co-operate" [116].

The current tests do not consider whether or not a person has Down's Syndrome, that therefore require more patience to have the test performed: it may take them longer to understand the tasks and it may be more difficult to keep them motivated in the test activities. For this reason it is important that children are assessed by people who have experience and interest in children with special needs [116] and to develop a test tool which can consider the real needs of this target of test takers.

Figure 6.15: The automated version of the McCormick Toy test - A portable digital speech screening system. Source: http://www.soundbytesolutions.co.uk/products/parrot

A company called Soundbyte Solutions[1] has produced two automated versions of the McCormick Toy Test (Fig. 6.15) known as the Parrot and the Phoenix[2], "portable digital speech screening systems", which provide speech stimuli via a speaker and uses an algorithm to vary the presentation level.

Few studies have examined the ability of individuals with DS to discriminate speech by means of technological systems based on speech recognition. Keller-Bell and Fox [114], for instance, compared the speech discrimination abilities of children with DS using a Windows-based laptop computer, a specifically designed software, and headphones: each child was presented with four sound stimuli and, at the

[1] http://www.soundbytesolutions.co.uk

[2] http://www.soundbytesolutions.co.uk/products/parrot/evaluation/

same time, four graphical boxes on the screen corresponding to each sound stimuli; each time the child heard a sound, one of the boxes would light-up. The participants were required to point the box of the different sound. The adult then indicated the childs response via the keyboard.

6.3.2 The ParrotGameDiscrimination Application

The need to develop a tool for the fine assessment of the sound/speech discrimination capabilities arises from the awareness that a difficulty at this level could be the reason of speech disorders of several atypical populations, among which Down's Syndrome, Specific Speech and Language Disorders, Hypoacusis. Psychologists, speech therapists, and educators require innovative tools besides paper-and-pencil tests for speech assessment. The introduction of multimedia tools offers an adequate answer to this need.

With this aim in mind, the CSC (*Centro di Sonologia Computazionale*) of the DEI (Department of Engineering Information) of the University of Padova, in collaboration with the DPSS (Department of Developmental and Socialization Psychology) of the University of Padova and this author, devised ParrotGameDiscrimination, a multimodal video game for mobile devices (e.g. tablets, pc), able to implement a method for the assessment of pupil's sound/speech discrimination skills.

The ParrotGameDiscrimination application provides test-takers with the opportunity to listen to pairs of sounds by means of a video game. The characters in the video game are, Momo, a little boy and Tiki, the parrot. Momo utters a sound and Tiki repeats it (Fig. 6.16).

Figure 6.16: The ParrotGameDiscrimination application - The characters in the video game are Momo, a little boy, and Tiki, the parrot. Momo utters a sound and Tiki repeats it.

The test-taker is prompted to assess whether Tiki repeated the sound/utterance in the same way. A "smiley" icon appears on the screen and the test-takers click on the smiley corresponding to their choice (Fig. 6.17). The sonorous stimuli reproduced by the two characters can be, the same, very similar, not very similar, or very different. This procedure allows assessing different levels of auditory discrimination, from gross global perception to very fine-grained recognition. The features of the sounds that are examined are the intensity, frequency, and timbre. The sounds that are reproduced are synthetic in order to mainly contain the assessed sound features. Recently, considering the complexity of these synthetic sounds, the database of sounds has been implemented with non-words [118].

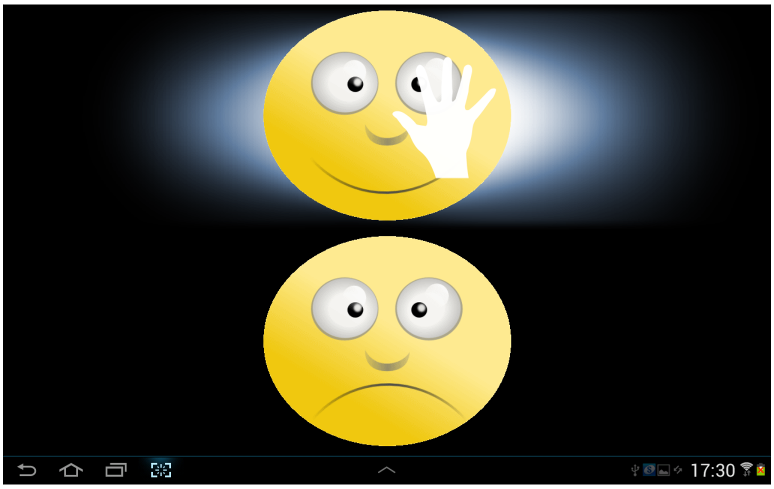

Figure 6.17: The PGD interaction - A "smiley" icon appears on the screen and the test-takers click on the smiley corresponding to their choice.

ParrotGameDiscrimination is made by several graphical components which provide operator and test-takers which several functions (Fig. 6.18). At the starting phase, the application presents a main screen which provides access the user to some software components, among these: Game Start, Visualize Results, Options. The first enables the user to start a test session and enter the test-taker's information (age, gender, name, and the ID of the session) (Fig. 6.19).

During the test, the application records the subject's responses and the reaction times. These data are stored into a database that is accessible through the Visualize Results function in the main menu. In Options, finally, the operator can modify some parameters: the pairs of sounds, their order, and some temporal parameters.

The ParrotGameDiscrimination application has been developed for the Android platform - 3.1 version Honeycomb - and, in particular it has been tested on a Samsung Galaxy Tablet 10.1. Java, XML e SQL

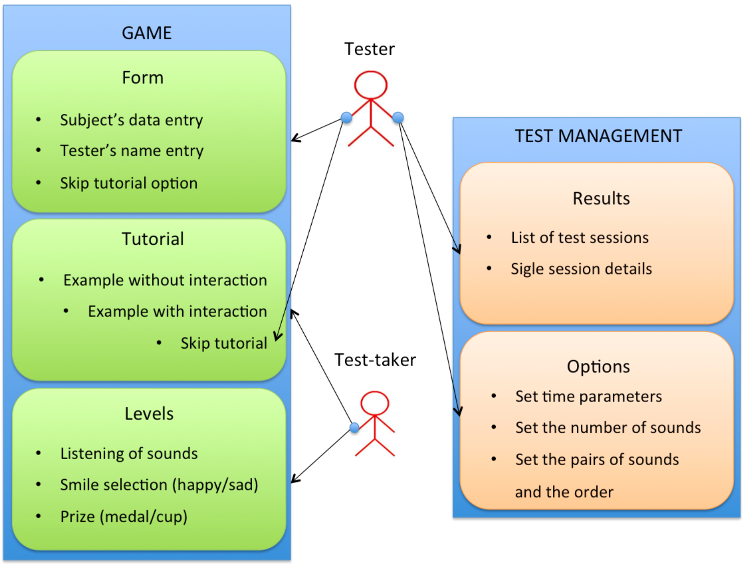

Figure 6.18: The ParrotGameDiscrimination functions - The system provides operators and test-takers with several functions. The scheme shows the possibilities of interaction for each user.

Figure 6.19: The test-taker's information - The PGD offers the possibility to enter the test-taker's information.

are the programming languages that have been employed for the development of the software; moreover, other tools and technologies, such as IDE Eclipse, DDMS, Adb, SQLite, and EGit, have been used.

6.3.3 Objectives of the System

Atypical language acquisition is recurrent and can be associated with: cognitive disability, social disadvantage, learning difficulties, or specific language impairments. Considering this complex context, the development of a) intervention strategies and b) fine assessment tools aimed at identifying the causes of language impairment is basic. The introduction of computer-based environments for assessment and intervention might provide support to this specific educational need.

The fine assessment of basic language skills is fundamental since many cases of language delay, phonologic impairment, Down's Syndrome, etc. are characterized by an impairment in speech production which might rely on difficulties in speech discrimination [119]. In language acquisition and literacy, the oral comprehension is very important as demonstrated by the Simple View of Reading model [64] attesting that reading comprehension relies on listening comprehension. This theory is also confirmed for pupils with Down's Syndrome [120]. However, it is always difficult to recognize whether a language disorder is due to a difficulty in discriminating acoustic characteristics of language or not. For this reason, ParrotGameDiscrimination has been designed.

The underlying idea is to use ParrotGameDiscrimination, a computer-based game, in order to obtain objective and reliable measures of speech perception skills. In particular, the ParrotGameDiscrimination, being a multisensory and interactive tool, conveys speech to a child through the auditory and visual channels. Specific aims of the ParrotGameDiscrimination are described as follows:

1. to measure the capability of children to discriminate, for each sound feature, the difference among a number of sound stimuli;
2. to improve and enrich the ludic experience of all the pupils and, at the same time, to promote the accessibility and the inclusion of the children with impairments;
3. to deepen the scientific research about the auditory processing of sound in specific populations of pupils who, due to auditory processing difficulties, encounter difficulties in speech production.

In order to reach all these objectives, ParrotGameDiscrimination, at the time of writing, is being experimented with 100 pre-school pupils aged 4 and 5. The experimentation is conducted by this author. Synthetic sounds and non-words [118] constitute the sound stimuli of the ParrotGameDiscrimination; moreover, a visual attention test [118] and the PPVT-R – Peabody Picture Vocabulary Test - Revised – ([121] and [122]) are administered.

7

General Discussion and Conclusions

The need to renew the teaching strategies by implementing innovative educational methods based on technological devices was evidenced within the Chapter 2. Popular, often utilized ICTs, though recognized as useful teaching tools, offer neither a high level of interactivity nor the experience of physical environment which are fundamental dimensions in children's cognitive development. For this reason, the Interactive and Multimodal Environments (IMEs), especially those embedded in physical spaces, have been identified as the best solution to satisfy the different requirements of a heterogeneous school population.

IMEs are systems that use sensors for data acquisition, real-time processing components for input data processing, and output components for reproduction of multimedia contents. IMEs are able to provide full-body interaction and multisensory experiences which enable the user to exploit the whole body and sounds to fulfill a task. Thus, learning is meaningful, since knowledge is obtained through a motivating sensory-motor experience.

Therefore, the main challenge of this doctoral research was to devise and develop IMEs that establish interaction between human-motion/speech features and physical spaces. It was argued that methodological demands about the analysis of human body interaction and multimodality for education are not entirely explored in the literature. This thesis project gradually approached these demands, by per-

forming a review of the literature (Chapters 2, 4, 5 and 6), developing new technological systems for teaching, and implementing preliminary studies and usability testing (Chapters 4, 5, and 6).
Chapter 3 fully presented the Stanza Logo-Motoria, an IME that involves the physical space, and the educational activities which can be implemented through it. Chapter 4 addressed the research method utilized to determine whether the use of the Stanza Logo-Motoria as a listening tool in ESL (English as a Second Language) classes has the hypothesized effect on the participants.
Starting from the experience of the Stanza Logo-Motoria, this thesis also illustrated the design, development, and evaluation of further IMEs, aimed at meeting specific educational objectives. These IMEs, fully presented in Chapter 5 and 6, were: SoundingARM (Sect. 5.1) and Memory Sonoro (Sect. 5.3), aimed at assisting people with visual impairments in orientation, mobility, and navigation skills acquisition, SoundRise (Sect. 6.2) which provides information on the acoustic properties of voice, and ParrotGameDiscrimination (Sect. 6.3) developed in order to enhance the assessment techniques of speech discrimination.
Each of the above IMEs was evaluated by means of various usability testing focused on measuring the capacity of the systems to meet the intended purpose. The results of the different evaluations highlighted the weak points, allowed to find solutions and improve the overall performance of the systems.

Today, the real challenge for educators is to organize learning environments, teaching practices, curricula, and resources that ensure effective learning. Given that learning is a very complex process, different cognitive perspectives have to be considered [123]. In this doctoral research project, Dual Coding [43], Enaction [37; 9; 35], Embodied Cognition [23; 59; 31; 34; 29], and Multimedia Learning [42] are the main theories wherein the employment of IMEs for learning is situated.
However, while these fields are ever evolving, researches have shown that significant increases in learning can be accomplished through the use of visual and verbal multimodal systems: "students engaged in learning that incorporates multimodal designs, on average, outperform students who learn using traditional approaches with single modes [124]."
The multiple representations of knowledge are now long considered as an effective way to facilitate learning [48]. Multiple representations usually involve the use of PowerPoint presentations as mini lectures, text synchronized with images, interactive diagrams, video presentations, audio explanations of concepts, and images. In these learning environments, the multimodal elements are merely presented as additional representations of information, thus the level of interactivity is very low.
Within an IME involving the physical space there are, instead, three main elements to address: the user, the environment, and other users. The IME, by stimulating more than one sense or interaction channel

(visual, auditory, but also haptic and olfactory), is able to create a wide range of interactions; among these, the user-to-user and the world-to-user interactions are the most important since they enable the user to experience and interact with others and the surrounding environment. IMEs, in order to manage multimodal interactions, need a hardware and software architecture able to "augment" the real world with multimedia contents; moreover, these systems have to be able to produce the appropriate stimuli over the various modalities in real time.
According to the foregoing assumptions, the research project documented in this thesis involved the design, development, and evaluation of IMEs able to provide learners with a meaningful learning by means of a multimodal and interactive experience. With this aim in mind, the Stanza Logo-Motoria and other specific technological systems have been developed in order to better deal with the complexity of the educational context.

The Stanza Logo-Motoria is the first IME implemented at school. By using standard hardware and simple strategies of mapping, this system helps both students and teachers in experiencing an interactive and multimodal way of learning and teaching. The Stanza Logo-Motoria, analyses the full-body movements and gestures of the users in real time, within a "sensorized" physical environment, and maps them onto real-time manipulation and processing of audio/visual content.
The Stanza Logo-Motoria offers greater access to knowledge and interaction with others and the environment enabling teachers and learners to discover the enactive approach of teaching/learning. Using the Stanza Logo-Motoria a great deal of educational activities, involving several school subjects, can be developed in collaboration with teachers. This system, by means of specific applications such as Fiaba Magica, also helps teachers to deal with pupils with severe disabilities who need to learn through alternative methodologies and tools specifically devised for them.
The Stanza Logo-Motoria is permanently installed in a Primary School since 2009. To date, both teachers and schoolchildren are using the system with interest and enthusiasm. For teachers this is due to the modular software architecture, simple strategies of mapping, and the possibility to easily customize the multimedia contents. Learners are still motivated by:

- the possibility to freely move in space activating audio-visual events (interactivity);
- the collaborative component of activities; there are a range of tasks that require a constant interchange with others and with multimedia contents using the modes of speech, image, movement, gesture and sound (multimodality);
- the opportunity to discover ever new and original multimedia contents (surprise effect).

The Stanza Logo-Motoria is suitable for the school environment also thanks to its easy implementation. The results of the two-year preliminary study confirmed, to a certain extent, the experimental hypothesis: with the Stanza Logo-Motoria, in Resonant Memory modality, learners are more likely to comprehend and remember what they listen to, compared to pupils using the language laboratory. Despite the small sample size, results indicating advantages of the Stanza Logo-Motoria – over the language laboratory – imply that there is an overall benefit in using this system. The statistical analysis showed that the differences between the experimental group and the control group are statistically significant.
In the future assessments, in order to gather more relevant data, more care has to be taken in:

1. test construction, making the difficulty and complexity of the listening test content appropriate to the test-takers' level;

2. the design and completion of the listening procedure, especially for the lessons of the control group; in fact, in the educational context many factors may easily influence the participants' attention and concentration, such as frequent interruptions of the lesson. From this point of view, the Stanza Logo-Motoria is effectively a more controlled environment.

During the PhD period, further IMEs were developed in order to reach more specific objectives/needs. The SoundingARM system (Chapter 5, Section 5.1), for instance, was devised to assist people with severe visual impairment in mobility, navigation, and orientation tasks. The evaluations of the system showed that, through SoundingARM, blind persons can build a meaningful cognitive map of the environment which can be used to navigate with more confidence and to explore the space without having to perform the locomotor/tactile exploration.
Memory Sonoro (Chapter 5, Section 5.3) is another system that uses sound to provide information which would otherwise be inaccessible to the blind. It is based on the very popular game "Memory", which consists in finding the pairs of identical cards/images. The pairs of pictures of the classical Memory game, in this case are replaced by pairs of markers (placed on cards) that the system recognizes and links to couples of sounds. With Memory Sonoro, therefore, the players must find the pairs of identical sounds. Moreover, the sounds are spatialized in real-time when the user shifts the tags on the horizontal or frontal plane. By means of Memory Sonoro users exploit tactile and auditory sensations in order to build mental images and to improve the sound discrimination skills. The evaluation of the system showed that Memory Sonoro is a useful tool to develop tactile-auditory perception integration. Memory Sonoro has a wide range of possible applications: it might be re-designed in order to trigger specific learning in children both with typical and atypical development. Moreover, it might improve or rehabilitate language, mathematical, and conceptual knowledge.

The SoundRise application (Chapter 6, Section 6.1) was designed, instead, with the purpose to provide pupils with an interactive and multimodal tool to learn the main features of the voice: pitch, intensity, timbre, and duration. The usability testing showed that by means of the SoundRise application pupils autonomously discover, interpret, and understand the information provided by the graphical feedback. Pupils have an active role since they control, by means of their voice, the graphical evolution of the feedback.

The ParrotGameDiscrimination (Chapter 6, Section 6.3) application provides pupils with the opportunity to listen to pairs of sounds by means of a video game. This application – designed for mobile devices (e.g. tablets, pc) – allows assessing different levels of auditory discrimination, from gross global perception to very fine-grained recognition, in order to recognize whether a language disorder is due to a difficulty in discriminating acoustic characteristics of language. Therefore, ParrotGameDiscrimination is a fine assessment tool designed in order to detect difficulties in speech discrimination. The features of the sounds examined are: intensity, frequency, and timbre. Synthetic sounds and non-words constitute the sound stimuli.

The educational possibilities that this type of technological environments might implement, if integrated with curriculum contents and skills, are endless; to date, there are few similar experiences, therefore, research on Interactive and Multimodal Environments for learning in real educational contexts is a field that has to be explored. However, the results of the research documented in this thesis suggest that, even if much work has still to be done, the premises are encouraging.

Following the approach taken in this thesis, future IMEs could be further developed and implemented in particular in the field of specific learning difficulties and special educational needs. Indeed, the educational context has to face with an increasing number of students with specific learning difficulties and disabilities which typically affect a student's motor skills, information processing, and memory. Teachers, in order to deal with this complex situation, require to implement "inclusive teaching" which means recognizing, accommodating, and meeting the learning needs of all the students. However, teachers' experience has demonstrated that adjustments made for students with disabilities can very often benefit all students.

A future development can be the creation of "dynamic environments" which allow the user to interact with them altering their structure and features. The goal will be to implement user-world and multi-user interactions which allow the user to interact with the real world in a much more engaging way. In this kind of environments learners can organize resources, manipulate information, and even create new content also in collaboration with others. Within dynamic environments students are not simply consumers of information, they become part of an active learning experience.

Another very interesting direction that can be followed in this research in the near future could be the development of techniques for the measurement of non-verbal social signals. The exploitation of non-verbal social interaction in interactive and multimodal systems can be very useful to support effective learning and co-creation.

References

[1] ITALIAN NATIONAL INSTITUTE OF STATISTICS. **The Integration of Disabled Students in Public and Private Primary and Lower Secondary Schools - School Year 2011-2012**. Technical report, Ministry of Education, Universities and Research (Miur), January 2013. 2

[2] UNICEF INNOCENTI RESEARCH CENTRE. **Children and Disability Intransition in Cee/Cis and Baltic States**. Technical report, UNICEF (United Nations Children's Fund), 2005. 2

[3] UNICEF INNOCENTI RESEARCH CENTRE. **Promoting the Rights of Children with Disabilities**. Technical report, UNICEF (United Nations Children's Fund), 2007. 2

[4] H. GARDNER. *Educazione e Sviluppo della Mente. Intelligenze Multiple e Apprendimento*. Centro Studi Erickson, 2005. 2, 18

[5] H. GARDNER. *Frames of Mind: The Theory of Multiple Intelligences*. Basic, 1983. 2, 18

[6] L. VYGOTSKY. *Thought and Language*. MIT Press, 1986. 2, 12

[7] D. H. JONASSEN. **Computers as Mindtools for Engaging Critical Thinking and Representing Knowledge**. In *EdTech99 Proceedings: Educational Technology Conference and Exhibition*, 1999. 2, 12

[8] S. PRICE AND Y. ROGERS. **Let's Get Physical: the Learning Benefits of Interacting in Digitally Augmented Physical Spaces**. *Computers and Education*, **43**:137–151, 2004. 2, 12, 13

[9] J. BRUNER. *Toward a Theory of Instruction*. Belknap Press of Harvard University Press, 1966. 2, 13, 15, 16, 102

[10] A. CAMURRI, S. CANAZZA, C. CANEPA, G. L. FORESTI, A. RODÀ, G. VOLPE, AND S. ZANOLLA. **The Stanza Logo-Motoria: an Interactive Environment for Learning And Communication**. In *Proceedings of SMC Conference 2010, Barcelona*, 2010. 3, 4, 28, 31, 56

[11] S. ZANOLLA, F. ROMANO, F. SCATTOLIN, A. RODÀ, S. CANAZZA, AND G. L. FORESTI. **When Sound Teaches**. In S. ZANOLLA, F. AVANZINI, S. CANAZZA, AND A. DE GÖTZEN, editors, *Proceedings of the SMC 2011 - 8th Sound and Music Computing Conference*, pages 64–69, 2011. 4, 5, 31, 35, 36

[12] S. ZANOLLA, S. CANAZZA, A. RODÀ, A. CAMURRI, AND G. VOLPE. **Entertaining Listening by means of the Stanza Logo-Motoria: an Interactive Multimodal Environment**. *Entertainment Computing Journal*, 2013. 4, 5, 31, 32, 35, 37, 50, 52

[13] L. COHEN, L. MANION, AND K. MORRISON. *Research Methods in Education*. Routhledge, 6 edition, 2007. 4, 41, 42, 43, 46, 48, 66

[14] D.T. CAMPBELL AND J.C. STANLEY. *Experimental and Quasi-Experimental Designs for Research on Teaching*, chapter 5. In Handbook of Research on Teaching. Chicago: Rand McNally, 1963. 4, 41

[15] N. SCATTOLIN, S. ZANOLLA, A. RODÀ, AND S. CANAZZA. **SoundingARM: Assisted Representation of a Map**. In *Proceedings of HAI Conference*, November 2012. 6, 56, 57, 58, 61, 62, 63, 64

[16] S. ZANOLLA, G. TEMPESTA, L. SNIDARO, AND S. CANAZZA. **Memory Sonoro: Realtà Aumentata Accessibile mediante Audio Direzionale**. In *Proceedings of the XIX CIM – Music Informatics Symposium*, 2012. 6, 7, 56, 70, 72, 73, 74, 76

[17] J. A BACHOROWSKI. **Vocal Expression and Perception of Emotion**. *Current Directions in Psychological Science*, **8**:53–57, 1999. 7, 82

[18] S. PATEL, K. R. SCHERER, E. BJÖRKNER, AND J. SUNDBERG. **Mapping Emotions into Acoustic Space: the Role of Voice Production**. *Journal of Biological Psychology*, **87**:93–98, 2011. 7, 82

[19] A. CAMURRI AND P. FERRENTINO. **Interactive Environments for Music and Multimedia. Special issue on Audio and Multimedia**. *ACM Multimedia Systems*, **7**(1):32–47, January 1999. 11, 12, 20

[20] J. NIELSEN. *Usability Engineering*. Morgan Kaufmann, 1994. 11

[21] A. CAMURRI, G. VOLPE, G. DE POLI, AND M. LEMAN. **Communicating Expressiveness and Affect in Multimodal Interactive Systems**. *IEEE Multimedia*, pages 45–55, January-March 2005. 12

[22] J. A JACKO AND A. SEARS. *The Human-Computer Interaction Handobook: Fundamentals, Evolving Technologies and Emerging Applications*. Lawrence Erlbaum Associates, 2003. 12, 16

[23] D. L. HOLTON. **Constructivism + Embodied Cognition = Enactivism: Theoretical and Practical Implications for Conceptual Change**. In *Proceedings of AERA Conference*, 2010. 12, 16, 102

[24] M. JAMES AND A. POLLARD. **Learning and Teaching in Primary Schools: Insight from TLRP (Primary Review Research Survey 2/4)**. Technical report, University of Cambridge, Faculty of Education, May 2008. 12

[25] S. ZANOLLA, S. CANAZZA, A. RODÀ, AND G. L. FORESTI. *A Learning Environment based on Movement and Sound Interaction (In press)*. Blue Herons, 2012. 12, 16, 40

[26] E. TSE, J. SCHÖNING, Y. ROGERS, C. SHEN, AND G. MORRISON. **Next Generation of HCI and Education: Workshop on UI Technologies and Educational Pedagogy**. In *CHI '10 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '10, pages 4509–4512, New York, NY, USA, 2010. ACM. 12

[27] S. FEINSTEIN. *The Praeger Handbook of Learning and the Brain*. Praeger, 2006. 13

[28] A. LUCIANI AND C. CADOZ. *Enaction and Enactive Interfaces: A Handbook of Terms*. Enactive Systems Books, 2007. 13, 16, 19

[29] M. WILSON. **Six Views of Embodied Cognition**. *Psychonomic Bulletin and Review*, **9**(4):645–636, 2002. 14, 102

[30] M. MERLEAU PONTY. *Phenomenology of Perception*. Routhledge Classics, 2005. 14

[31] GEORGE LAKOFF AND MARK JOHNSON. *Philosophy in the Flesh: The Embodied Mind and its Challenge to Western Thought*. Basic Books, 1999. 14, 102

[32] G. RIZZOLATTI AND L. VOZZA. *Nella mente degli altri. Neuroni specchio e comportamento sociale.* Zanichelli, 2008. 14, 15

[33] E. KOHLER, C. KEYSERS, M.A. UMILTÀ, L. FOGASSI, V. GALLESE, AND G. RIZZOLATTI. **Hearing Sounds, Understanding Actions: Action Representation in Mirror Neurons**. *Science*, **297**(5582):846, 848, August 2002. 14, 15

[34] M. LEMAN. *Embodied Music Cognition and Mediation Technology*. The MIT Press, 2007. 15, 31, 102

[35] H. MARTURANA AND F. VARELA. *Autopoiesis and Cognition: The realization of the Living.* Reidl, 1980. 15, 16, 102

[36] M. CLAYTON, R. SAGER, AND U. WILL. **In Time with the Music: the Concept of Entrainment and its Significance for Ethnomusicology.** In *ESEM Counterpoint*, **1**, pages 1–82, 2004. 15

[37] J. BRUNER. *Processes of Cognitive Growth: Infancy.* Clark University Press, Worcester, MA, 1968. 15, 16, 31, 102

[38] J. STEUER. **Defining Virtual Reality: Dimensions Determining Telepresence.** *Journal of Communications*, **42**(4):73–93, Autumn 1992. 16

[39] I. E. SUTHERLAND. **Sketchpad: a Man-Machine Graphical Communication System**. *Technical Report*, (574), September 2003. 16

[40] Y. ROGERS AND M. SCAIFE. **How Can Interactive Multimedia Facilitate Learning?** *Intelligence and Multimodality in Multimedia Interfaces: Research and Applications. AAAI*, 1998. 16

[41] R. E. MAYER. *Cambridge Handbook of Multimedia Learning*, chapter 3, pages 31–48. Cambridge University Press, 2005. 17

[42] R. E. MAYER AND R. MORENO. **A Cognitive Theory of Multimedia Learning: Implications for Design Principles**. In *Proceedings of ACM SIGCHI Conference on Human Factors in Computing Systems*, April 1998. 17, 102

[43] J. M. CLARK AND A. PAIVIO. **Dual Coding Theory and Education**. *Educational Psychology Review*, **3**(3):149–210, October 2006. 17, 102

[44] L. J. NAJJAR. **Multimedia Information and Learning**. *Journal of Educational Multimedia and Hypermedia*, **5**(2):129–150, 1996. 18

[45] D. BIRCHFIELD, THOMAS CIUFO, H. THORNBURG, AND W. SAVENYE. **Sound and Interaction for K-12 Mediated Education**. In *International Computer Music Conference Proceedings*, 2006. 18

[46] L. L. SCARLATOS. **TICLE: Using Multimedia Multimodal Guidance to Enhance Learning**. Number 140, pages 85–103. Elsevier, 2002. 19

[47] B.R. JOHNSON. **Virtuality and Place**. In *ACADIA*, 2002. 20

[48] R. MORENO AND R. MAYER. **Interactive Multimodal Learning Environments**. *Special Issue on Interactive Learning Environments: Contemporary Issues and Trends. Educational Psychology Review*, pages 309–326, 2007. 20, 102

[49] J. DAVIS F. BAIRD A. BOBICK, S. INTILLE, C. PINHANEZ, L. CAMPBELL, Y. IVANOV, A. SCHUTTE, AND A. WILSON. **The KidsRoom: A Perceptually-Based Interactive and Immersive Story Environment**. In *Presence: Teleoperators and Virtual Environments*, **8**, pages 367–391, 1999. 20, 22

[50] B. B. BEDERSON AND A. DRUIN. *Computer Augmented Environments: Physical Spaces to Enrich Our Lives*, **5**, chapter 2, pages 37–66. Ablex Press, 1995. 20, 28

[51] M. KRUEGER, T. GIONFRIDDO, AND K. HINRICHSEN. **Videoplace - An Artificial Reality. Human Factors in Computing Systems**. *ACM press*, 1985. 21

[52] M. KRUEGER. **VIDEOPLACE: a Report from the ARTIFICIAL REALITY Laboratory**. *Leonardo*, **18**(3):145–151, 1985. 21

[53] M. KRUEGER. *Responsive Environments. The New Media Reader*. The MIT Press, 1977. 21

[54] P. MAES, T. DARRELL, B. BLUMBERG, AND A. PENTLAND. **The ALIVE System: Wireless, Full-Body Interaction with Autonomous Agents**. *Special Issue on Multimedia and Multisensory Virtual Worlds, ACM Multimedia Systems*, Spring 1996. 21

[55] G. DAVENPORT AND G. FRIEDLANDER. *Interactive Transformational Environments: Wheel of Life*, chapter 1, pages 1–25. MIT Press, Cambridge, 1995. 21

[56] A. CAMURRI AND A. COGLIO. **An Architecture for Emotional Agents**. *IEEE Multimedia*, **5**(4):24–33, Oct-Dec 1998. 21, 22

[57] YVONNE ROGERS, SARA PRICE, ERIC HARRIS, TED PHELPS, MIA UNDERWOOD, DANIELLE WILDE, HILARY SMITH, MARK THOMPSON MARK J. WEAL, AND DANIUS T. MICHAELIDES. **Learning Through Digitally-Augmented Physical Experiences: Reflections on the Ambient Wood Project**. Technical report, 2002. 23, 24

[58] J. A. FAILS, A. DRUIN, M. L. GUHA, G. CHIPMAN, S. SIMMS, AND W. CHURAMAN. **Child's Play: a Comparison of Desktop and Physical Interactive Environments**. In *Proceedings of the 2005 Conference on Interaction Design and Children*, pages 48–55, New York, USA, 2005. ACM. 24

[59] M. C. JOHNSON-GLENBERG, D. BIRCHFIELD, P. SAVVIDES, AND C. MEGOWAN-ROMANOWICZ. *Semi-virtual Embodied Learning - Real World STEM Assessment*, pages 225–241. Sense Publications, 2010. 24, 25, 102

[60] S. ZANOLLA, S. CANAZZA, A. RODÀ, AND G. L. FORESTI. **Learning by Means of an Interactive Multimodal Environment**. In ALAIPO AND AINCI, editors, *Proceedings of SETECEC Conference*, March 2012. 25, 34

[61] A. CAMURRI AND T. B. MOESLUND. *Visual Gesture Recognition. From Motion Tracking to Expressive Gesture*. Appears as chp. 10 in the book Musical Gestures. Sound, Movement, and Meaning. Rolf Inge Godøy and Marc Leman (Eds.). Published by Routledge, ISBN: 9780415998871, 2010. 29

[62] T. HORPRASERT, D. HARWOOD, AND L. DAVIS. *A Robust Background Subtraction and Shadow Detection*, **1**. In 4th ACCV, Taipei, Taiwan, 2000. 31

[63] L. AMICO. *La Stanza Logo-Motoria. Un ambiente Multimodale Interattivo per l'Insegnamento a Bambini in Situazione di Multi-Disabilità*. Master's thesis, University of Padova - Department of Information Engineering, october 2012. 32, 33

[64] P. B. GOUGH. **How Children Learn to Read and Why They Fail**. *Annals of Dyslexia*, **46**:3–20, 1996. 34, 100

[65] L. BAINES. *A Teacher's Guide to Multisensory Learning - Improving Literacy by Engaging the Senses*. Association for Supervision and Curriculum Development (ASCD), 2008. 34

[66] W. MATTHEW. **Open Sound Control: an Enabling Technology for Musical Networking**. *Organised Sound*, **10**(3):193–200, 2005. 36

[67] M. CALABRESE. *Fiaba Magica: Sviluppo e Sperimentazione di un Ambiente Interattivo Multimodale per l'Attuazione di Percorsi Didattici Inclusivi*. Master's thesis, University of Padova - Department of Information Engineering, December 2012. 36, 38

[68] S. HOWELL AND L. KESTER-DODGSON. *Treetops Plus Class Book*, **3**. Oxford University Press, 2010. 45, 46, 47

[69] S. HOWELL AND L. KESTER-DODGSON. *Treetops Plus Class Book*, **4**. Oxford University Press, 2010. 45, 46, 47

[70] L. COTTINI. *Fare Ricerca nella Scuola dell'Autonomia*. Mursia, 2002. 46

[71] M. H. COEN. **Design Principles for Intelligent Environments**. In *Proceedings of the 1998 National Conference on Artificial Intelligence*, 1998. 56

[72] C. WILLIAMS. **Unintentional Intrusive Participation in Multimedia Interactive Environments**. In *Proceedings of 7th ICDVRAT with ArtAbilitation*. ICDVRAT/University of Reading, 2008. 56

[73] R. GEHLHAAR. **SOUND=SPACE: an Interactive Musical Environment**. *Contemporary Music Review*, **6**(1):59–72, 1991. 56

[74] R. GEHLHAAR, L. M. GIRAO, AND P.M. RODRIGUEZ. **Cadaremi - An Educational Interactive Music Game**. *ICDVRAT with Art Abilitation*, 2008. 56

[75] F. GHEDINI, H. FASTE, M. CARROZZINO, AND M. BERGAMASCO. **Passages - A 3D Artistic Interface for Child Rehabilitation and Special Needs**. In *Proceedings of 7th ICDVRAT with ArtAbilitation*. ICDVRAT/University of Reading, UK, 2008. 56

[76] S. HASSELBLAD, E. PETERSSON, AND T. BROOKS. **Interactivity in Work with Disabled**. In *Proceedings ArtAbilitation*, 2006. 56

[77] A. BOBICK, S. INTILLE, J. DAVIS, F. BAIRD, C. PINHANEZ, L. CAMPBELL, Y. IVANOV, A. SCHÜTTE, AND A. WILSON. *Design Decisions for Interactive Environments: Evaluating the KidsRoom*. 1998. 56

[78] R. RAMANATHAN. *Combining Egocentric and Exocentric Views: Enhancing the Virtual Worlds Interface*. 1999. 58

[79] C. THINUS-BLANC AND F. GAUNET. **Representation of Space in Blind Persons: Vision as a Spatial Sense?** *Psychological Bulletin*, **121**(1):20–42, 1997. 58

[80] Y. HATWELL. *Psicologia Cognitiva della Cecità Precoce*. Biblioteca Italiana per i Ciechi "Regina Margherita" - ONLUS, 2010. 58, 64

[81] Y. HATWELL. *Psychologie Cognitive de la Cécité Précoce*. Dunod, Paris, 2003. 58, 64

[82] E. HILL AND P. PONDER. *Orientation And Mobility Techniques: a Guide for the Practitioner*. New York: American Foundation for the Blind, 1976. 58

[83] J.M. LOOMIS, R.G. GOLLEDGE, AND R.L. KLATZKY. **Navigation System for the Blind: Auditory Display Modes and Guidance**. *Presence*, **7**(2):193–203, April 1998. 59

[84] M. HERSH. *Assistive Technology for Visually Impaired and Blind People*. Springer Verlag, 2008. 59, 60

[85] Y. SONNENBLICK. **An Indoor Navigation System for Blind Individuals**. In *Proceedings of the 13th annual Conference on Technology and Persons with Disabilities*, 1998. 60

[86] L. RAN, S. HELAL, AND STEVE MOORE. **Drishti: an Integrated Indoor/Outdoor Blind Navigation System And Service**. In *IEEE International Conference on Pervasive Computing and Communications*, 2004. 60

[87] C. MAGNUSSON, M. MOLINA, K. RASSMUS-GRÖHN, AND D. SZYMCZAK. **Pointing for Non-Visual Orientation and Navigation**. In *Proceeding of NordiCHI, Reykjavik, Iceland*, October 2010. 60

[88] S. MAU, N. A. MELCHIOR, M. MAKATCHEV, AND A. STEINFELD. *BlindAid: An Electronic Travel Aid for the Blind*. 2008. 60

[89] A. HUB, J. DIEPSTRATEN, AND T. ERTL. **Design and Development of an Indoor Navigation and Object Identification System for the Blind**. *SIGACCESS Access. Comput.*, (77-78):147–152, 2004. 60

[90] M. C. WANET AND C. VERAART. **Processing of Auditory Information by the Blind in Spatial Localization Tasks**. *Perception and Psychophysics*, (38):91–96, 1985. 64

[91] D. MARABESE. *Sviluppo e Sperimentazione di un Auditory Display per la Deambulazione di Non Vedenti*. Master's thesis, University of Padova - Department of Information Engineering, October 2012. 65

[92] P. KELLER AND C. STEVENS. **Meaning From Environmental Sounds: Types of Signal–Referent Relations and Their Effect on Recognizing Auditory Icons**. *Journal of Experimental Psychology*, **10**(1):3–12, 2004. 68

[93] S. FRAUENBERG AND M. NOISTERING. **3D Audio Interfaces for the Blind**. In *Proceedings of the 2003 International Conference on Auditory Display*, July 2003. 70, 71

[94] V. DELIĆ AND N. V. SEDLAR. *Stereo Presentation and Binaural Localization in a Memory Game for the Visually Impaired*, **5967**, pages 354–363. Springer-Verlag Berlin Heidelberg, 2010. 71, 72

[95] J. SÁNCHEZ AND M. SÁENZ. **Three-Dimensional Virtual Environments for Blind Children**. *Cyberpsychology and Behavior*, **9**(2):200–206, 2006. 71, 72

[96] M. LUMBRERAS AND J. SÁNCHEZ. **3D Aural Interactive Hyperstories for Blind Children**. *The International Journal of Virtual Reality*, **3**(4):18–26, 1998. 71

[97] J. H. SÁNCHEZ AND H.E. FLORES. **AudioMath: Blind Children Learning Mathematics through Audio**. In *Proceedings of the 5th Intl Conf. Disability, Virtual Reality and Assoc. Tech.*, pages 183–189. ICDVRAT/University of Reading, 2004. 71

[98] T. WESTIN. **Game Accessibility Case Study: Terraformers - a Real-Time 3D Graphic Game**. In *Proceedings of the 5th Intl Conf. Disability, Virtual Reality and Assoc. Tech.*, pages 95–100. ICDVRAT/University of Reading, 2004. 72

[99] M. OBRIST, F. FÖRSTER, D. WURHOFER, M.TSCHELIGI, AND J.HOFSTÄTTER. **Evaluating First Experiences with an Educational Computer Game: A Multi-Method Approach**. *Interaction Design and Architecture(s) Journal - IxDandA*, (11-12):26–36, 2011. 78

[100] D. A. NORMAN. *The Design of Everyday Things*. The MIT Press, 1998. 79

[101] O. BÄLTER, O. ENGWALL, A. M. ÖSTER, AND H. SIDENBLADH-KJELLSTRÖM. **Wizard-of-Oz Test of ARTUR - a Computer-Based Speech Training System with Articulation Correction**. In *Proc. 7th Int. ACM SIGACESS Conf. Computers and accessibility (ASSETS'05)*, pages 36–43, 2005. 81

[102] H. Fell, C. Cress, J. MacAuslanm, L. Ferrier, G. Sterup, and A. Heinrich. **Visi-Babble for Reinforcement of Early Vocalization**. In *Proceedings of ASHA Conference*, 2004. 81

[103] J. Hailpern, K. Karahalios, J. Halle, L. DeThorne, and M.K. Coletto. **Visualizations: Speech, Language and Autistic Spectrum Disorder**. In *CHI '08 extended abstracts on Human factors in computing systems (CHI EA '08)*, 2008. 81

[104] M. Randon. *Soundrise: Sviluppo e Validazione di un'Applicazione Multimodale Interattiva per la Didattica Basata sull'analisi di Feature Vocali*. Master's thesis, University of Padova, July 2012. 81, 85

[105] S. Giusto. *Soundrise: Studio e Progettazione di un'Applicazione Multimodale Interattiva per la Didattica Basata sull'analisi di Feature Vocali*. Master's thesis, University of Padova, July 2012. 81, 85, 87, 89

[106] R. Godwin-Jones. **Emerging Technologies Speech Tools and Technologies**. *Language Learning and Technology*, **133**:4–11, 2009. 82

[107] S.M. Williams, D. Nix, and P. Fairweather. **Using Speech Recognition Technology to Enhance Literacy Instruction for Emerging Readers**. In B. Fishman and S. O'Connor-Divelbiss, editors, *Fourth International Conference of the Learning Sciences*, pages 115–120, 2000. 82, 83

[108] K. Vicsi and Á. Váry. **Distinctive Training Methods and Evaluation of a Multilingual, Multimodal Speech Training System**. In *Proc. 4th Intl Conf. Disability, Virtual Reality and Assoc. Tech.* ICDVRAT/University of Reading, 2002. 82

[109] S. Harada, J.O. Wobbrok, and J.A. Landay. **VoiceDraw: A Hands-Free Voice-Driven Drawing Application for People with Motor Impairments**. In *ASSETS '07*, October 2007. 83, 84

[110] D.M. Howard, G. F. Welch, J. Brereton, E. Himonides, M. DeCosta, J. Williams, and A.W. Howard. **WinSingad: a Real-Time Display for the Singing Studio**. *Logoped Phoniatr Vocol*, **29**:135–144, June 2004. 84

[111] A. de Götzen, R. Marogna, and F. Avanzini. **The Voice Painter**. In *Proc. Int. Conf. on Enactive Interfaces*, 2008. 84, 85

[112] OFSTED. **Music in Schools: Wider still, and Wider - Quality and Inequality in Music Education 2008-11**. Technical report, Office for Standard in Education, Children's Services and Skills (Ofsted), March 2012. 85

[113] J. SIMNER, J. WARD, M. LANZ, A. JANSARI, K. NOONAN, L. GLOVER, AND D. A. OAKLEY. **Non-Random Associations of Graphemes to Colours in Synaesthetic and Non-Synaesthetic Populations**. *Cognitive Neuropsychology*, **22**(8):1069–1085, 2005. 88

[114] Y. KELLER-BELL AND R. A. FOX. **A Preliminary Study of Speech Discrimination in Youth with Down Syndrome**. *Clinical Linguistics and Phonetics*, **21**(4):305–317, April 2007. 95, 96

[115] L. K. BENNETTS AND M. C. FLYNN. **Improving the Classroom Listening Skills of Children with Down Syndrome by using Sound-Field Amplification**. *Down Syndrome Research and Practice*, **8**(1):19–24, 2002. 95

[116] J. DENNIS. **Hearing Problems in People With Down's Syndrome. Notes for Parents and Carers**, 2001. 95, 96

[117] B. MCCORMICK. *The Medical Practioner's Guide to Paediatric Audiology*. Cambridge University press, 1995. 95

[118] P. S. BISIACCHI, M. CENDRON, M. GUGLIOTTA, P. E. TRESSOLDI, AND C. VIO. *BVN 5-11 Batteria di Valutazione Neuropsicologica per l'Età Evolutiva*. Test e Strumenti di Valutazione Psicologica Educativa. Erickson, 2005. 98, 100

[119] B. DODD AND L. THOMPSON. **Speech Disorder in Children with Down's Syndrome**. *Journal of Intellectual Disabilty Research*, **45**(4):308–316, August 2001. 100

[120] M. ROCH AND M. C. LEVORATO. **Simple View of Reading in Down's Syndrome: the Role of Listening Comprehension and Reading Skills**. *International Journal of Language and Communication Disorders*, **44**(2):206–223, March-April 2009. 100

[121] G.J. ROBERTSON, L.M. DUNN, J.L. EISENBERG, AND L.M. DUNN. *PPVT: Peabody Picture Vocabulary Test - Revised*. American Guidance Service, 1981. 100

[122] G. STELLA, C. PIZZOLI, AND P.E. TRESSOLDI. *Peabody - Test di Vocabolario Recettivo*. Omega, 2000. 100

[123] M. STEWARD. **Learning Through Research: an Introduction to the Main Theories of Learning**. *JMU Learning and Teaching Press*, **4**(1):6–14, 2004. 102

[124] C. FADEL AND C. LEMKE. **Multimodal Learning Through Media: What the Research Says**. Technical report, Cisco Systems, Inc., 2008. 102

List of Figures