The Sounding Object

Edited by

Davide Rocchesso and Federico Fontana

Edizioni di Mondo Estremo

The Sounding Object

Davide Rocchesso, Editor Università di Verona, Italy

Federico Fontana, Editor Università di Padova and Università di Verona, Italy

Federico Avanzini Università di Padova, Italy

Nicola Bernardini Conservatorio di Padova, Italy

Eoin Brazil University of Limerick, Ireland

Roberto Bresin Kungl Tekniska Högskolan, Stockholm, Sweden

Roberto Burro Università di Udine and Università di Padova, Italy

Sofia Dahl Kungl Tekniska Högskolan, Stockholm, Sweden

Mikael Fernström University of Limerick, Ireland

Kjetil Falkenberg Hansen Kungl Tekniska Högskolan, Stockholm, Sweden

Massimo Grassi Università di Udine and Università di Padova, Italy

Bruno Giordano Università di Udine and Università di Padova, Italy

Mark Marshall University of Limerick, Ireland

Breege Moynihan University of Limerick, Ireland

Laura Ottaviani Università di Verona, Italy

Matthias Rath Università di Verona, Italy

Giovanni Bruno Vicario Università di Udine, Italy

The work described in this book was supported by the European Commission Future and Emergent Technologies collaborative R&D program under contract IST-2000-25287 (project SOb - the Sounding Object).
The SOb project web site is http://www.soundobject.org
The book web site is http://www.soundobject.org/SObBook

ISBN 88-901126-0-3

Cover Design: Claudia Calvaresi

Produced from LaTeX text sources and PostScript images.
Online distributed in Portable Document Format.
Printed and bound in Italy by PHASAR Srl, Firenze.

Contents

Preface

In the last decade of the twentieth century, new research trends emerged in the study of sound and its applications: (i) In the sound synthesis community, physical models have been increasingly used and appreciated for their natural dynamics and ease of control; (ii) In the psychoacoustic community, several researchers started looking at the perceived qualities of sound sources and events, thus embracing an ecological attitude that shifts from classical signal-based studies; (iii) Researchers in music acoustics developed sophisticated models to render the expressive nuances of music performance and started to extend their interest to everyday, human-driven acoustic phenomena; (iv) Sound started to find its place in the broad field of human-computer interaction, with the emergence of new auditory interfaces and displays.

At the turn of the century, the four mentioned research areas were separate and largely independent from each other, but we had the feeling that something good could be done by asking these various people to work together. The opportunity arose when the proactive initiative on the Disappearing Computer was launched by the European Commission in the year 2000. We just had to find a good argument for a new research project on sound, and this was: If computers tend to disappear as cognitive artefacts, their visual displays becoming tiny or huge, we should exploit the auditory channel for effective machine-human communication via non-speech structured sound. The idea, formalized in the proposal for a project called SOb - the Sounding Object, wouldn't have turned into actual research without the enthusiastic support of: (i) the European Commission, especially the project officer Loretta Anania, (ii) the researchers of the Universities of Verona, Udine (led by Giovanni Bruno Vicario), Limerick (led by Mikael Fernström) and KTH - Stockholm (led by Roberto Bresin), (iii) all researchers from other institutions and projects who joined us in public initiatives and research activities. The reader can find extensive documentation on what these people have done and are continuing to do in the project web site at http://www.soundobject.org.

This book collects the most interesting scientific results of the SOb project. Even though it is a compendium of the research done, we would like to use it to push and steer our future activities. We have just begun to appreciate the benefits of having computer scientists doing experimental psychology, psychologists doing computer models, and acousticians doing interfaces. This hybridization of communities takes time to develop, and this book may contribute to speed up the process by letting us know each other better.

The book is also an example of free publishing. It is printed on paper and available on the web in source form, for free. We believe that scientific tools, data, and literature should be accessible and verifiable by all. Conversely, we notice that contemporary publishing policies are too often driven by other, non-scientific interests. Of course, publishing scientific results in a book like this is not an alternative to peer-reviewed journal publication, but it is rather a complement of usual scientific practices. In fact, by opening our work to everybody we implicitly assume that any reader can be a reviewer and contribute either to the acceptance and development of the ideas here expressed or to their rejection.

Many people contributed in a way or another to the progress of the SOb project and to the realization of this book. Let us just mention the project reviewers Bill Gaver and Phil Ellis, whose criticism has always been stimulating, and our friend Nicola Bernardini, who is constantly keeping us busy with his flow of ideas, proposals, and ready-made computer tricks.

Davide Rocchesso and Federico Fontana, Venezia 16th April 2003

Foreword

This book is a much-coveted and nurtured breakthrough in several domains. First and foremost, of course, it is one of the most complete and detailed scientific presentations to date of ill-defined entities such as “sounding objects”, tackling hirsute aspects like precise physical modeling, object behavior and perceptual features.

The difficulty in defining these sounding object entities come from the fact that they belong both to our everyday auditory display and to that very specific attitude that we observe when we listen to music. Indeed, music is “whatever we intend to hear as music” — as Luciano Berio used to say [18], and sounding objects constitute the basic building bricks of both music and auditory landscape. The studies presented in this book do not try to separate these aspects; as a matter of fact, in several of them music — along with its panoply of specific features such as form, expressivity, intention, etc. — appears side by side to strict scientific data treatments (for ex. while analyzing rhythmic features or DJ scratching techniques). The extent to which this interesting cross-breeding is practiced in this book is, once again, quite unique.

Another peculiarity derives directly from this mingling of music and sound. This book is the distillation of the work carried out in a two-year long project of the same name which has brought together highly skilled specialists with very different backgrounds to work on the same subject from widely different perspectives. While this is certainly not unique in recent years, the close interaction among researchers to draw a clear picture out of diversified results is not easily found elsewhere.

Finally, the very same publishing of this book has its characteristic features: this book is both available in its paper form and in a non-proprietary electronic format over the Internet. Furthermore, all the sources of this book are published under the GNU Free Documentation License. We firmly believe that this publishing form will strongly enhance not only its diffusion, but also the very same use that scholars and students will make of it: while all authors have done their best to produce the most compelling book on sounding objects, they have also given out the possibility to make it better.

Nicola Bernardini

Chapter 1

Everyday listening: an annotated bibliography

Bruno L. Giordano
Università di Udine – Faculty of Education
Udine, Italy
Università di Padova – Department of General Psychology
Padova, Italy
bruno.giordano@unipd.it

There is an immense world of sounds that have been scarcely considered by traditional research in auditory perception, the world of everyday sounds. We should systematize it, understand its own regularities and laws.

Many of the sounds we encounter every day are generated by physical events that involve an interaction between objects (a coin dropped on the floor, water drips falling on the sink) or a change of the properties of single objects (a bursting balloon). Are listeners able to recover the properties of these physical events on the basis of auditory information alone? What is the nature of the information used to recover these features? These questions were originally raised inside the ecological approach to perception, firstly extended to the study of the auditory modality by Vanderveer [240]. Gaver [92, 97] outlined a taxonomy of environmental sounds, in order to “entice other explorers” into this domain. His taxonomy is based on the simple assertion that sounds are generated by an “interaction of materials”, and it is inside this map that we should place the pieces of experimental evidence collected so far, synthesized in this chapter.

Research in sound source recognition requires methodological tools, of course. A formalization of the research design in this field is found in [153]. The object under study can be described at three levels: the physical, acoustical and perceptual ones. All the pairwise investigations between these three levels of description should receive our attention. Analysis of the relationship between the perceptual and physical levels tells us if the investigated feature of the physical event is properly recognized or scaled by listeners. Analysis of the relationship between the physical and acoustical levels tells us which properties of the acoustical signal differentiate between categories or levels of the physical feature of interest. Analysis of the relationship between the acoustical and perceptual levels tells us whether the acoustical information outlined at the previous stage is effectively used by participants to scale or categorize the physical feature of interest or whether other acoustical features influence listeners responses. This last analysis should be validated by manipulation of the acoustical properties found significant in explaining participants responses.

A corollary to Li et al. [153] formalization is that we should assess whether recognition of a particular feature of the sound source is possible despite variations in extraneous physical features. In other words we should make use of what we define as *perturbation variables*, defined as those variable features of the sound source, extraneous from the one whose recognition is investigated. For example Kunkler-Peck and Turvey [142] studied scaling of the dimensions of struck plates upon variations in their dimensions and material. In this case material is a perturbation variable.

Finally we would like to make the reader aware of one more issue. Research on sound source recognition could demonstrate several misalignments between the physical features of the sound source and the recognized ones, especially if perturbation variables are extensively used. The misalignment between physical and perceptual levels is a well known lesson in the visual perception field, and there will be no surprise to empirically demonstrate this in the auditory domain. In this case research on sound source recognition will turn as research on the *apparent features of the sound source*. In this case we will have to start thinking about this type of research as an alternative way for addressing a problem which is still widely unknown: the problem of the quality of auditory percepts. As the problem of quality in audition has been already addressed by research on timbre perception (see [171, 113] for a review), we would like to conclude this introduction using a metaphor in debt with these studies. The world of sounds can be conceived as organized within an multidimensional space. Classical research on auditory perception have investigated this space along dimensions such as pitch, loudness, duration, timbral brightness, attack hardness and so on. Research on sound source recognition envisions the existence of a set of new dimensions to investigate this multidimensional space, criteria based on the features of the sound source, be them coherent with the physical reality or apparent.

1.1 1979

S. J. Lederman. *Auditory texture perception* [148]

Lederman compared the effectiveness of tactile and auditory information in judging the roughness of a surface (i.e., the *texture*). Roughness of aluminum plates was manipulated by varying the distance between adjacent grooves of fixed width, or by varying the width of the grooves. The subjects' task was to rate the roughness of the surface numerically. In the auditory condition, participants were presented the sounds generated by the experimenter who moved his fingertips along the grooved plate. In the remaining conditions, subjects were asked to move their fingertips onto the plate: In the tactile condition they wore cotton plugs and earphones while touching the plate; in the auditory-plus-tactile condition they were able to ear the sound they generated when touching the plate.

Roughness estimates were not different between the auditory-plus-tactile and tactile conditions, but differed in the auditory condition. In other words when both kinds of information were present, the tactile one played the strongest role in determining experimental performance. Roughness estimates were shown to increase as both the distance between grooves and the width of the grooves decreased. Additionally roughness estimates increased as the force exerted by the finger over the surface increased, and the speed of the finger movement decreased. The effect of the force on roughness estimates in the auditory condition was however not constant across subjects. A speculative discussion concerning the relative role of pitch and loudness in determining the estimates is provided by the author, although no acoustical analyses of the experimental stimuli are provided.

N. J. Vanderveer. *Ecological Acoustics: Human perception of environmental sounds* [240]

Vanderveer's dissertation applies for the first time the ecological approach to auditory perception. Empirical evidence was collected using three different methodologies. In the first study a tape containing 30 environmental sounds (e.g., crumpling paper, jingling coins, wood sawing) was played continuously to two groups of subjects. They were asked to write down what they heard. Descriptions of the event were given by subjects based on the sound features, rather than the acoustical properties of the stimuli. However, subjects were found to refer to abstract perceptual qualities when they did not recognize the sound source. Accuracy was fairly good in all cases. Descriptions focused mainly on the actions causing the sound events, rather than the objects' properties. Confusions were frequent among stimuli sharing common temporal patterns, when such patterns were hypothesized to carry information about the actions.

In a second experiment subjects were asked to classify 20 environmental sounds into a number of different categories, based on perceived similarity. Again, clustering seemed to be based on the similarity of the temporal patterns. Results collected using the same classification procedure were confirmed by another experiment rating similarity, that was conducted using a subset of the stimuli adopted in the classification experiment.

1.2 1984

W. H. Warren and R. R. Verbrugge. *Auditory perception of breaking and bouncing events: a case study in ecological acoustics* [251]

According to the terminology adopted by those embracing the ecological approach to perception, we distinguish between two classes of invariants (i.e., higher-order acoustical properties) that specify the sound generation event. Structural invariants specify the objects properties, whereas transformational invariants specify the way they change. Warren and Verbrugge investigated the nature of the structural invariants that allow identification of breaking and bouncing events. By conducting a physical analysis of these two kinds of events, the authors hypothesized that the nature of these invariants was essentially temporal, the static spectral properties having no role in the identification of breaking and bouncing events. Experimental stimuli were generated by dropping on the floor one of three different glass objects from different heights, so that for each object a bouncing event and a breaking one was recorded. Once the ability of the participants to correctly identify these two types of events was assessed using the original stimuli, two further experiments were conducted making use of synthetic stimuli. The bouncing event was synthesized by superimposing four damped, quasi-periodic pulse trains, each one generated by recording one of four different bouncing glass tokens. These four sequences exhibited the same damping coefficient. The breaking event was synthesized by superimposing the same pulse trains, this time using a different damping coefficient for each of them (in the second experiment the breaking stimuli were preceded by a 50 ms noise burst of the original breaking sound). The identification performance was extremely accurate in all cases, despite the strong simplifications of the spectral and temporal profile of the acoustical signal. Therefore the transformational invariants for bouncing were identified to be a single damped quasi-periodic sequence of pulses, whereas those for breaking were identified to be a multiple damped, quasi-periodic sequence of pulses.

1.3 1987

B. H. Repp. *The sound of two hands clapping: an exploratory study* [200]

Speech perception has been classified by Liberman and Mattingly as perception of phonetic gestures [154]. This theory has been given the name of motor theory of speech perception. Repp's work extends this theoretical approach to the investigation of non-speech communicative sound: claps. In particular, Repp hypothesized the subjects' ability to recognize the size and the configuration of clapping hands by auditory information. The recognition of the hand size was also put in relation with the gender recognition of the clapper, given that male have in general bigger hands than females. Several clapping sounds were recorded from different clappers. In the first experiment the recognition of the clapper' gender and his/her hand size was investigated indirectly, as participants were asked to recognize the clapper's identity. Recognition was not good, although the listeners' performance in the identification of their own claps was much better. Gender recognition was barely above chance. Gender identification appeared to be guided by misconceptions: faster, higher-pitched and fainter claps were judged to be produced by females and vice-versa. In the second experiment, subjects had to recognize the configuration of the clapping hands. Subjects were found to be able to recover correctly the hand configuration from sound. Although the hand configuration was significant in determining the clapping sound spectrum, nevertheless the best predictor of performance was found to be the clapping rate, the spectral variables having only a secondary role during the recognition task.

W. H. Warren, E. E. Kim, and R. Husney. ***The way the ball bounces: visual and auditory perception of elasticity and control of the bounce pass*** **[250]**

Warren et al. studied the perception of elasticity in bouncing balls, in both the visual and auditory domains. In experiment 1 subjects were requested to bounce a ball off the floor to a constant target height. Five balls with different elasticities were used, and subjects were exposed to different kinds of information concerning the elasticity of the ball before executing the task (in the auditory condition they heard the sound of a ball dropped on the floor, whereas in the auditory plus visual condition they saw and heard the same ball bouncing on the floor). Results showed that prior exposure to both visual and auditory information about the ball's elasticity did not lead to a different performance level, compared to when only auditory information was provided. In experiment 5 subjects had to rate the elasticity of a bouncing ball, simulated with a white circle bouncing on a white line at the bottom of the screen, and with a 40 ms, 190 Hz tone that was presented at visual impact time to simulate the bouncing sound. The rated bounciness was the same when it was judged from auditory or visual information only. In all cases, subjects appeared to judge elasticity by evaluating a single inter-bouncing period, rather than the ratio between successive periods.

1.4 1988

W. W. Gaver. ***Everyday listening and auditory icons*** **[92]**

In this section two empirical investigations reported by Gaver in his dissertation are summarized. In the first investigation participants were presented 17 different environmental sounds, generated by the interaction between solid objects and/or liquids. They were asked to describe what they heard, possibly providing the highest level of detail. Impacts were always identified: subjects were able to extract information concerning the object material, and possibly its size and hollowness. The accuracy in the identification depended on the specificity of the details. Crumpling can sounds were often confused with multiple impacts, probably because of the common physical nature of the two events. Interestingly, when crumpling sounds were generated using paper, confusion between crumpling and group of impacts was rare. Liquid sounds were correctly identified in all cases. Results gathered using more complex sounds again revealed a high performance level. For example, electric razor sounds were always recognized as being generated by a machine. Walking sounds were always correctly identified, with three subjects correctly recognizing sounds as being generated by a person walking upstairs, rather than downstairs.

Two experiments on real struck bar sounds were then performed. In the first experiment subjects had to categorize the material of struck metal and wooden bars of different lengths. Performances between 96 and 95% correct were observed. In the second experiment subjects had to estimate the length of the same bars from sound. A control group of subjects was, instead, asked to estimate the pitch of the same set of stimuli. Length ratings were almost a linear function of the physical length, the type of material being non-significant. For length estimates an interaction between material and length was found, so that the shortest and the longest metal bars were estimated to be shorter than wooden bars of equal length. The psychophysical functions for pitch scaling were much different for the two materials, the metal bar sounds having a higher pitch than the wooden ones. Similar results were found for all the tasks using synthetic bar sounds, except for the length rating. In this case, modeled changes of the bar length were associated to smaller variations in the estimated length compared to those observed using real sounds.

R. Wildes and W. Richards. *Recovering material properties from sound* [253]

The purpose of the authors was to find an acoustical parameter that could characterize material type uniquely, i.e. despite variations in objects features such as size or shape. Materials can be characterized using the coefficient of internal friction $tan\phi$, which is a measure of anelasticity (in ascending order of $tan\phi$ we have rubber, wood, glass, and steel). In the acoustical domain the coefficient of internal friction was found to be measurable using both the quality factor Q^{-1} and the decay time of vibration t_e, this latter measured as the time required for amplitude to decrease to $1/e$ of its initial value. For increasing $tan\phi$ we have an increase in Q^{-1}, and in t_e.

1.5 1990

D. J. Freed. *Auditory correlates of perceived mallet hardness for a set of recorded percussive events* [82]

Freed's study aims to measure an attack-related timbral dimension using a sound source-oriented judgment scale: hardness. Stimuli were generated by percussing four cooking pans, having variable diameter, with six mallets of variable hardness. Mallet hardness ratings were found to be independent of the pan size, thus revealing the subjects' ability to judge the properties of the percussor independently of the properties of the sounding object. The main goal of this study was to derive a psychophysical function by mallet hardness ratings, based on the properties of the acoustical signal. Preliminary experiments pointed out that significant information about mallet hardness was contained in the first 300 ms of the signals. For this reason, the acoustical analyses focused on this portion of the signals. Four acoustical indexes were measured: average spectral level, spectral level slope (i.e., rate of change in spectral level, a measure of damping), average spectral centroid, and spectral centroid time weighted average (TWA). These acoustical indexes were used as predictors in a multiple regression analysis. Together, they accounted for 75% of the variance of the ratings.

1.6 1991

X. Li, R. J. Logan, and R. E. Pastore. *Perception of acoustic source characteristics: Walking sounds* [153]

Li et al. studied gender recognition in walking sounds. Walking sounds of seven females and seven males were recorded. Subjects were asked to categorize the gender by a four-step walking sequence. Results show recognition levels well above chance. Several anthropometric measures were collected on the walkers. Male and female walkers were found to differ in height, weight and shoe size. Spectral and duration analyses were performed on the recorded walking excerpts. Duration analysis indicated that female and male walkers differed with respect to the relative duration of the stance and swing phases, but not with respect to the walking speed. Nonetheless, judged masculinity was significantly correlated with the latter of these two variables, but not with the former. Several spectral measures were derived from the experimental stimuli: spectral centroid, skewness, kurtosis, spectral mode, average spectral level, and low and high spectral slopes. Two components where then derived by applying a principal components analysis on the spectral predictors. These components were used as predictors for both physical and judged gender. Male walkers in general were characterized by having a lower spectral centroid, mode and high frequency energy than females, and by higher values of skewness, kurtosis and low-frequency slope. The same tendencies were found when the two components were used as predictors for the judged gender.

Results gathered from the analysis of the relationship existing between the acoustical and the perceptual levels were then tested in another experiment. Stimuli were generated by manipulating the spectral mode of the two most ambiguous walking excerpts (also the spectral slopes were altered, but the manipulation of this feature was not completely independent of the manipulation of the spectral mode). Consistently with previous analyses, the probability of choosing the response "male" was found to decrease with increasing spectral mode. A final experiment showed that the judged gender could be altered by making a walker wear shoes of the opposite gender.

1.7 1993

W. W. Gaver. *What in the world do we hear? An ecological approach to auditory event perception* [97]

In the everyday life we recognize events and sound sources rather than sounds. This listening attitude has been defined by Gaver as "everyday listening", as opposed to "musical listening" where the perceptual attributes are those concerned by traditional research in audition. Despite the behavioral relevance of non-musical and non-speech sounds, empirical researches on them are missing. Research on everyday sounds focuses on the study of new perceptual features and dimensions, those concerning the sound source. Analyzing how the sound source features structure the acoustical signal is thus necessary, to find a set of candidate dimensions. This analysis, however, does not tell us which of these dimensions are relevant to everyday listening. For this purpose it is thus necessary to use protocol studies. The map of everyday sounds compiled by Gaver is based on both the knowledge about how a sound source structures the acoustical signal, as well as on protocol studies data. The most important distinction is found between solid, liquid and aerodynamic sounds, as protocol studies showed that these macro-classes are seldom confused each other. Then each of these classes is divided based on the type of interaction between materials. For example, sounds generated by vibrating solids are divided in rolling, scraping, impact and deformation sounds. These classes are "basic level sound-producing events". Each of them make different sound source properties evident.

The next level contains three types of complex events: those defined by a "temporal patterning" of basic events (e.g., bouncing is given by a specific temporal pattern of impacts); "compounds", given by the overlap of different basic level events; "hybrid events", given by the interaction between different types of basic materials (i.e., solids, liquids and gasses). Each of these complex events should potentially yield the same sound source properties, made available by the component basic events plus other properties (e.g., bouncing events may provide us informations concerning the symmetry of the bouncing object).

W. W. Gaver. ***How do we hear in the world? Explorations in ecological acoustics*** **[95]**

A study on everyday listening should investigate both the relevant perceptual dimensions of the sound generation events (i.e., what we hear) and the acoustical information through which we gather information about the events (i.e., how we hear). In timbre perception the study of relevant acoustical information can be based upon the so-called analysis and synthesis method [203]. This methodology looks for relevant information by progressively simplifying the acoustical structure of the signals investigated, until only the acoustical properties, whose further simplification would lead to relevant timbral changes, are retained. Likewise, everyday-sound synthesis algorithms are developed after the analysis of both the acoustical and physical event, complemented by the perceptual validation which has been made possible by the synthesis stage. Several algorithms are presented for the synthesis of impact, scraping, dripping, breaking/bouncing/spilling, and machine sounds. A final discussion highlights on some of the methodological issues that are connected to the validation of synthesis models.

1.8 1997

S. Lakatos, S. McAdams, and R. Caussé. ***The representation of auditory source characteristics: simple geometric form*** **[145]**

Lakatos et al. studied the listeners' ability to discriminate the shape of steel and wooden bars, as specified by the ratio between their height and width (H/W). Sounds were generated striking the bars with a mallet. All the stimuli were equalized in loudness. A cross-modal matching task was performed by the subjects, in which they had to indicate which one of two possible sequences of figures (representing the two bars with different H/W ratios) corresponded to the sequence heard. Stimuli generated by percussing steel and wooden bars were tested in different sessions.

Subjects who did not reach a 75% correct criterion were excluded from further analyses (8.3% of the subjects did not reach that criterion for the steel bars, and 16.6% did not reach it for the wooden bars). The correct scores, converted in the appropriate way, were analyzed using MDS techniques. A two-dimensional solution was derived for the data related to steel bars. The coordinates labeling the first dimension were highly correlated with the H/W ratio, and with the frequency ratio of the transverse bending modes. The coordinates labeling the second dimension highly correlated with the spectral centroid. A cluster analysis of the same data set revealed a gross distinction between thick and thin bars (blocks vs. plates). Data gathered on wooden bars sounds led to a one-dimensional MDS solution, with the same correlation properties exhibited by the first dimension of the MDS solution derived in the case of steel bars.

R. A. Lutfi and E. L. Oh. ***Auditory discrimination of material changes in a struck-clamped bar*** **[163]**

Lutfi and Oh studied the discrimination of material in synthetic struck clamped bar sounds. Stimuli were synthesized by varying the parameters of bar elasticity and density toward characteristic values which characterize iron, silver, steel, copper, glass, crystal, quartz, and aluminum, respectively. Perturbations were applied either to all the frequency components (lawful covariation) or independently to each component (independent perturbation). On half of the trials participants had to tell which one of two presented stimuli was an iron sound, silver, steel, and copper being the alternatives. On the other half of the trials the target was glass, and the alternatives were crystal, quartz, and aluminum. Participants were given feedback on the correctness of their response after each trial. Performances were analyzed in terms of weights given to three different acoustical parameters: frequency, decay, and amplitude. The data revealed that discrimination was mainly based on frequency in all conditions, the amplitude and decay rate having only a secondary role.

1.9 1998

C. Carello, K. L. Anderson, and A. J. Kunkler-Peck. ***Perception of object length by sound*** **[49]**

Carello et al. investigated the recognition of the length of wood rods dropped on the floor. In two experiments, the former focusing on longer rods, subjects judged the perceived length by adjusting the distance of a visible surface put in front of them. Subjects were able to scale the rod length consistently. The physical length was found to correlate strongly with the estimated length ($r = 0.95$ in both cases), although the latter experiment showed a greater compression of the length estimates (slope of the linear regression function equal to 0.78 in the former experiment, and to 0.44 in the latter experiment). An analysis of the relationships existing between the acoustical and perceptual levels was carried on using three acoustical features: signal duration, amplitude, and spectral centroid. Apart from the logarithmic amplitude in the latter experiment, none of the considered acoustical variables predicted the length estimates better than the actual length. Length estimates were finally explained by means of a kinematic analysis of the falling rods. The results of this analysis show potential analogies between the auditory and the tactile domain.

V. Roussarie, S. McAdams, and A. Chaigne. ***Perceptual analysis of vibrating bars synthesized with a physical model*** **[209]**

Roussarie et al. used the MDS methodology to study a set of stimuli, which were synthesized by a physical model of a vibrating bar. The bar density and the damping factor were used as synthesis parameters. All stimuli were equalized in loudness and fundamental frequency. Subjects had to rate, according to an analogical scale, the perceived similarity between paired stimuli. They received no information concerning the nature of the stimuli. Similarity ratings were analyzed with MDS algorithms. A two-dimensional solution was found: The first dimension was well correlated with a power function of the damping factor (a parameter of the physical model) and with a linear combination of the logarithm of the decay time of the amplitude envelope and of the spectral centroid. The second dimension correlated with the bar densities and with the frequency of the second component, that was taken as an estimate of perceived pitch.

1.10 2000

P. A. Cabe and J. B. Pittenger. ***Human sensitivity to acoustic information from vessel filling*** **[46]**

Cabe and Pittenger studied vessel filling by listeners' judgments during different tasks. The first experiment assessed the listeners' ability to distinguish filling events from similar events. Stimuli were generated by pouring water into an open tube. The apparatus was designed so to increase (filling), decrease (emptying), or leaving constant the water level inside the tube during pouring. Subjects were asked to categorize stimuli using these three event-categories. Identification accuracy values ranged from 65% to 87%, depending on the type of event. In experiment 2, subjects were asked to fill the vessel up to the brim or to the drinking level. In the former condition only auditory information was available, whereas in the latter one subjects could use all the available perceptual information (visual, tactile, auditory etc.). Results showed a better performance in the latter condition. Nonetheless, in the auditory condition filling levels were close to the maximum possible level.

In the third experiment, blind and blindfolded subjects were asked to fill to brim vessels of different sizes, and with different water flow velocities. Overall performance was accurate. The vessel size, the flow velocity and their mutual interaction influenced the error (computed as the height of the unfilled portion of the vessel), in a way that it was maximum for the smallest vessel filled with the fastest flow velocity, and minimum for the largest vessel. Conversely, when the error was computed as a percentage of the unfilled vessel height with respect to the vessel total height, then the opposite profile was found. Furthermore, no significant differences between blind and blindfolded participants were found.

Patterns of change specifying the time remaining before the end of an event have been defined as τ [149]. A similar variable may be used by listeners to judge the time remaining for a vessel to be filled to the brim. The final experiment tested this hypothesis. Blindfolded participants were asked to judge the time required for a vessel to be filled to the brim using the sounds generated by filling a vessel to three different levels, with three different flow rates. Responses consistently varied with the filling level, and the estimated filling time was strongly correlated with the actual filling time, this revealing an effective use of the τ variable during the execution of the task.

R. L. Klatzky, D. K. Pai, and E. P. Krotkov. *Perception of material from contact sounds* [137]

Klatzky et al. investigated material discrimination in stimuli with variable frequency and decay modulus τ_d. In the first two experiments subjects had to judge on a continuous scale the perceived difference in the material of an object. Stimuli had the same values of frequency and τ_d, but in the second experiment they were equalized by overall energy. As results did not differ significantly in the two experiments, it could be concluded that intensity is not relevant in the judgment of material difference. Experiments 3 and 4 were conducted on the same set of stimuli used in experiment 2. In the former subjects had to judge the difference in the perceived length of the objects, in the latter they had to categorize the material of the objects using four response alternatives: rubber, wood, glass and steel. Results indicated that judgments of material difference and of length difference were significantly influenced by both τ_d and frequency, even though the contribution of the decay parameter to length difference was smaller than that to material difference. An effect of both these variables was found in a categorization task: for lower decay factors steel and glass were chosen over rubber and plexiglass. Glass and wood were chosen for higher frequencies than steel and plexiglass.

A. J. Kunkler-Peck and M. T. Turvey. *Hearing shape* [142]

Kunkler-Peck and Turvey investigated shape recognition from impact sounds. In the first experiment stimuli were generated by striking steel plates of constant area and variable height/width with a steel pendulum. Participants had to reproduce the height and width of the plates by adjusting the position of several bars within a response apparatus. Although dimensions were underestimated, the subjects' performance revealed scaling of a definite impression of the height and width of plates (i.e., definite scaling). Simple regression models were computed using the plates' dimensions or modal frequencies as predictors. Both predictors were highly efficient in motivating the subjects' performance, as regression models were associated to r^2 coefficients which were higher or equal to 0.95. In the second experiment subjects were asked to scale the dimensions of constant-area and variable-height/width plates made of steel, plexiglass and wood. The type of material was discovered to simply modulate the subjects' estimates (i.e., it was associated to a simple additive effect without altering the estimated H/W ratio). Again, the scaling performance was well justified by the theoretical modal frequencies of the plates.

The remaining two experiments were designed to address shape recognition directly. In the third experiment stimuli were generated by striking a triangular, a circular and a rectangular steel plate (the area was kept constant). Shape was correctly classified at a level well above chance. In the last experiment stimuli were generated by striking circular, rectangular and triangular plates made of steel, wood and plexiglass. Subjects were asked to categorize the material as well as the shape. The material was almost perfectly classified, and shape was correctly classified at a level well above chance. A curious tendency of subjects to associate specific geometrical shapes to specific material types was reported (wood with circle, steel with triangle, plexiglass with rectangle).

S. Lakatos. *A common perceptual space for harmonic and percussive timbres* [144]

This research provides a direct empirical link between timbre perception and sound source recognition. Stimuli produced using sounds from musical instruments, either producing harmonic tones and percussive sounds, were investigated. All stimuli were equalized in pitch and loudness. Eighteen musicians and sixteen non-musicians were asked to rate analogically the timbral similarity of paired stimuli. In three separate sessions subjects had to rate the harmonic set, the percussive set, and a mixed set including harmonic as well as percussive sounds. Responses were analyzed using MDS techniques, as well as clustering procedures. In all cases a musical training did not appear to determine strong differences in the response profiles. For the harmonic set, the MDS analysis revealed a clustering based on the mode of excitation (impulsive vs. continuous). In fact, the first dimension of the MDS space correlated strongly with the logarithmic rise time. Consistently with that, the principal division among stimuli, as computed by the clustering procedure, divided impulsive from continuous tones. Minor divisions were based on the nature of the proximal stimulus, mainly depending on the spectral centroid proximity rather than on features of the source properties. The second dimension of the MDS solution was highly correlated with the spectral centroid.

A three-dimensional MDS solution was derived for the percussive tones. The first dimension was correlated with log rise time, the second with spectral centroid. The third dimension was associated with "timbral richness". Interestingly, the cluster analysis revealed that the stimuli grouped depending on the features of the physical source. For example membranophones clustered together, as well as instruments with wood cavities or instruments made with metallic plates. A two-dimensional MDS solution was derived for the combined set. Dimensions correlated, respectively, with log rise time and with spectral centroid. String instruments were quite overlapped with bar and tubular percussion instruments, probably because of similarities existing in their physical structure. The cluster solution again revealed a grouping based on the similarities in the physical structure of the instruments: Bars, strings and struck tubes were clustered together, as well as wind instruments, drums, and metal plate instruments.

R. A. Lutfi. *Auditory detection of hollowness* [162]

Lutfi investigated the recognition of hollowness using stimuli synthesized according to the equation describing the motion of a clamped bar. The equation parameters of density and elasticity were chosen in order to model iron, wood and aluminum bars. During each trial subjects were presented the sound of a hollow and of a solid bar. They were asked to tell which one of the two stimuli had been generated by striking a hollow bar. The inner radius of the hollow bar was chosen in order to keep a performance level between 70% and 90% correct. The bar length was randomly chosen from a normal distribution with a specific mean (10 cm for iron and aluminum, 25 cm for wood bars) and a standard deviation of 0.5 cm. Feedback on the correctness of the response was given after each trial. An analysis of the decision weights revealed two strategies used by different listeners to perform the task. One group adopted a decision strategy based on partial frequencies and decay times, that allowed optimal discrimination between hollow and solid bars. The other group adopted a decision strategy based only on frequency. The effect on performance of a limited processing resolution of the acoustical cues was then analyzed. In this way it was shown that an optimal decision strategy yielded, at best, a small advantage over the decision rule based on frequency.

Chapter 2

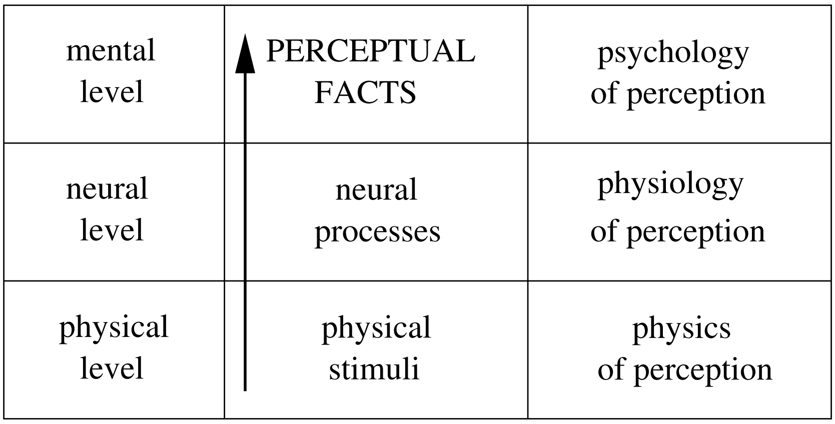
Prolegomena to the perceptual study of sounds

Giovanni Bruno Vicario
Università di Udine – Faculty of Education
Udine, Italy
vicario.gb@for.uniud.it

2.1 The perceptual study of sounds

The study of objects and events in visual field developed through interminable discussions on the nature of perceptual processes and on the methods that can provide well founded knowledge about the matter. Profound theoretical divergences are still on stage, namely between the Helmholtzean view (which cognitivistic, ecological, computer based and neurophysiological approaches after all refer to) and the Gestalt view (supporting the irreducibility of mental facts, like perception, to any machinery inside the brain). In visual field takes also place the question of any distinction between objects and events (some ones assert, some others deny it), since the perception of events (changes in quality, in position, in number) brings to the most intractable problem: that of psychological time.

I think that some echoes of that discussions may be useful for the students of perception of sounds, supposing that in all sensory modalities (vision, audition, touch *etc.*) the strategies of the perceptual system are the same. Be sure, in auditory field the presence of time is even more apparent, because the unceasingly flow of stimulation sometimes takes the form of an evolving event, and sometimes gives rise to perception or to representation of the source of the stimulus (for example in active touch).



mental level	PERCEPTUAL FACTS	psychology of perception
neural level	neural processes	physiology of perception
physical level	physical stimuli	physics of perception

Figure 2.1: The three levels of perception.

2.2 Perception and psychophysics

Perception is a process that goes through three levels of reality and may be analyzed at three levels, as the scheme of Figure 2.1 shows. By means of this scheme, I try to synthesize some facts well known to perceptionists:

(a) the description of perceptual facts in the sole terms of the stimuli is useless—this is the so called *stimulus error*;

(b) the description of perceptual facts in the sole terms of neural processes is misleading—this is the so called *experience error*;

(c) perceptual facts are linked to neural processes, and – as a rule – to physical stimuli;

(d) there are robust methods to describe perceptual facts at the physical and neural levels;

(e) in order to describe perceptual facts at the perceptual level, there is only a way: to ask experimental subjects for the description [155].

Now, verbal responses of subjects – as well as motor responses subsequent to the tasks – are biased by an endless list of "errors" that begins with individual thresholds, goes through misinterpretations of the task and judgement uncertainties and eventually reaches expectations and motivations. Psychophysics is the discipline that tries to avoid all the more important sources of error, by mean of methods that assign reliable contours to unobservable objects like percepts. In a sense, the experimental subject is a measuring instrument, and we know that a measuring instrument is as much reliable as measures only the quantity for which it has been devised (for instance, the standard meter bar measures the length, and not the temperature that can increase or diminish its length). The development of psychophysics – lasting now since 150 years – shows that the effort to force the experimental subject to report only the facts we want to know is still in progress, by means of a crowd of methods more and more sophisticated. Yet we did not reach methods unconditionally valid.

Let us consider for instance the problem of previous experience. The perception and the recognition of certain sounds takes an obvious advantage from the knowledge the subject owes of that sounds, but there is no mean to weight that knowledge in a certain subject. That is because (a) any attempt to measure the quantity of exposition of the subjects to an already known sound is subdued to their ability of recovering all their past experiences, and (b) the impact of previous experiences is not always linked to their number – see, for example, the case of aversive behaviour, that is elicited by just one harmful episode. Training in experimental sessions, or learning by feedback are surely useful stratagems, but they leave the problem untouched, because their influence is restricted to the experience one can have during the training or the feedback procedures.

To conclude, psychophysical measurements have to be handled with a lot of care. It is sufficient to bring slight variations in procedure – both in tasks or in judgement scales – to turn over results. Someone will point out that this is a commonplace in science, and I agree.

2.3 Geographical and behavioural environment

If we want to attain the point of view of the psychologist, we have to start with the distinction between geographical and behavioural environment [138]. The geographical environment is the physical world, the source of all external stimuli. Living beings do not react to all energy exchanges having place in the physical world, but only to the ones useful or perilous for their survival. It is trivial to refer to ultraviolet rays that are real stimuli for bees, but are inexistent for us, or to ultrasounds, that are real stimuli for dogs, but are inexistent for us. The point is that the behaviour is not determined by potential stimuli physically present in the geographical environment, but by the sole stimuli that a filtering system (sensory system) allows to enter the living being.

Besides, we face also a rearrangement and disguise of stimuli that are in many ways disrespectful of physical reality. An unceasing research in vision demonstrated that:

1. we see objects that do not exist (anomalous surfaces);
2. we do not see objects that really exist (masking, mimicry, camouflage);
3. we see objects whose existence is impossible (Penrose, Escher);
4. we see two different objects in the same set of stimuli (ambiguous figures);
5. we see the same object from different points of view without moving the object or the observer (Necker's cube, reversible figures);
6. we see objects which exhibit properties different from the physical ones (visual illusions).

For the details on the matter, see Vicario [243]. To sum up, our visual environment is in many ways different from the physical, but we act on the basis of what *we see*, and not on the basis of what *there is*. Even some engineers seem to share this point of view [233].

As a consequence, if our field of researches regards the interaction between sound sources and humans, we have to ascertain (1) what are the acoustic stimuli that are relevant for human behaviour, and (2) what are the properties of the auditory (perceptual, subjective) world.

The task (1) is realized by the physiology of the ear and of the brain if we shed the fact that thresholds undergo considerable variations due to attention, expectations and other strictly psychological factors. The task (2) is to be realized, and in accord with studies in visual domain we are almost sure that the dimensions of auditory, behavioural world are other than the dimensions of the acoustic, physical world. Let us now consider the difference between physical and phenomenal behaviour in the auditory field. A rough exam of facts tells us that:

1. we hear sounds that do not exist (the missing fundamental);

2. we do not hear sounds that really exist (masking);

3. we hear two different objects with the same set of stimuli (reversible rhythms);

4. we hear sounds as produced by physical sources inexistent in the everyday environment (electronic music, and any spectral manipulation of real sounds);

5. as far as I can understand, this class is void, since time is an one-dimensional set;

6. there are a lot of auditory illusions (in tonal domain, see Deutsch [63]).

Since we act on the basis of what *we hear*, and not on the basis of what *there is*, we can conclude that the distinction between physical environment and behavioural environment holds even in auditory field. This fact introduces the problem of the recognition of the source of a sound.

2.4 Hearing the source

There is a problem in the statement that “we hear the source of sound”. That could be true in the natural environment: when a violin is playing in front of us, we could say that “we hear a violin”. But since we hear a violin also in the reproduction of its sound by means of a hi-fi apparatus, there is to understand why we do not say that “we hear the cone of the loudspeaker”. In other words, we do not “hear” the source, but we “represent” it. The obvious explanation that the cone of loudspeaker reproduces very well the acoustic stimulus proceeding from the violin does not work, since *in fact* we fail to identify the actual physical source of the stimulus (the cone of the loudspeaker) and we represent a source (a violin) that is *beyond* the hi-fi apparatus.

The meaning of the fact becomes quite clear when we listen to a vinile disc that, because of a damage between grooves, at a certain point reproduces the same set of stimuli. When we listen, for example, to a reproduction of a piece of music, we do not hear the loudspeaker or the machine behind it: we perceive the musician, or the orchestra. At the first iteration caused by the damage, we continue to hear the musician or the orchestra, as performing again the same set of tones to obtain a special musical effect (as occurs in final bars of some symphonies). Yet at the third or fourth iteration, we realize that something goes wrong: the musician or the orchestra dissolve, and we have the representation of an apparatus that does not work. There must be something unclear, in the passage from the acoustic stimulus to the verbal report of the listener.

In the visual domain the fact is well known as “pictorial perception” [101]. There are cases in which we perceive only the physical setting that is the tinted surfaces, for example a white wall, or a painting by Mondrian. There are other cases by which we do not see the tinted surfaces, but the represented object, for example an outlined apple, or a photograph, or even a person portrayed by Antonello da Messina. Finally, there are also cases of portrayed objects, architectural elements or even painted landscapes, by which mean we get the impression to be in presence of the “reality” itself: I refer to the *trompe l’oeil* phenomenon. In the auditory domain the problem has not yet been discussed, but it is apparent that we have also perception of a sound without the representation of the source (for example, when we refer to strange noises or timbres of tones, or to cries of unknown animals). That leads to the problem of recognition.

2.5 On recognition

The universal use of the term “recognition” deserves some comments.

1. First of all, it cannot stay in place of “perception”, because we have perception without recognition: we mentioned above the case of sounds that “we don’t know”, but the hearing of enigmatic or puzzling noises is a common experience. On the other hand, we have “recognition” without perception: the photocell of an elevator “recognizes” the presence of a person between the closing doors and stops them, but the electronic apparatus cannot be credited with perception.

2. In the second place, we can have recognition at various levels about the same sound. For example, once the stimulus is produced, I can recognize a sound and not a smell; I can recognize a noise and not a tone; I can recognize the noise as that of a car and not of a motorcycle; I can recognize that the noise is that of a VW and not of a BMW; I can recognize that the car is on the point of stopping, and is not passing away; I can recognize that the way of stopping is that of my second son, and not of my wife, and so on. It is plain that a term that does not identify an object or a process, or even a field of possibilities, is useless.

3. There are “false recognitions”: laying in my bed, I can recognize the noise of the rain, but when getting up to close the window, I realize that the noise was that of the leaves of the trembling poplars ruffled by the wind. Notice that false recognitions are not phenomenally different from true recognitions: they are always “true”, and they become “false” after the comparison with other sense data, by mean of a process that is neither perception nor recognition.

4. There are “ambiguous recognitions”: the same voice may in turn appear as belonging to a man or to a woman, to a woman or to a child.

5. There are “changing recognitions”, in the sense that a noise issued by the same source, when presented again to the perceiver, gives rise to different “recognitions”: the rehearsal of an unknown noise leads to changing verbal reports that reflect a change in individuating the source of the noise. Otherwise, in a reiterate succession of tones to which one tone is added at each iteration, the listener can individuate different already known melodies.

6. It is trivial to note that the same acoustic stimulus gives rise to different “recognitions” when some clones of that stimulus went before and after it: a single shot leads to the recognition of a pistol, whereas a sequence of shots leads to the recognition of a machinegun. Besides, the uncertainty of the temporal limits within which there is integration along stimuli (2, 3, or 4 shots?) in view of the recognition of their source, does not allow to identify the stimulus on which the process of “recognition” is exerted.

7. Sometimes the “recognized” source is immaterial: the acceleration or the slowing down of a rhythm has no physical counterpart, because the intervals between beats are time, that is not a stimulus, but a void container of events. The sounds produced by the way of walking of an individual is a temporal pattern that is not a physical object, and nevertheless the individual (this is the way of walking of my wife) or its expressive content (she is in a hurry) is identified.

8. Again, sounds exhibit expressive contents that, in the case of individuals, are surely immaterial (where is the hurry of my wife?), and, in the case of objects, are quite imaginary (this clock has gone mad).

9. There are recognitions that occur in real time, for instance when we listen to a continuous tone. Yet there are recognitions of very brief sounds that are exerted not on the physical signal, or on the representation of the source, but on the memory trace of the perceived sound. And we do not know what changes percepts undergo when transformed in memory traces (that could be ascertained for visual objects that are at least stationary, but is surely uncertain for auditory events, since the memory trace of an event is not an event) nor what changes traces undergo when recovered to enter the field of judgement, that is working memory or consciousness.

Be sure, we can put aside all those problems, and continue our research on sounds in artisan way, claiming that they are matter of psychological fancies. But if the psychologist has got involved, his duty is to stress the precariousness of some technical terms and the risk of hasty generalizations.

2.6 On physical models

Working for the “Sounding Object” project, I have been acquainted with two strategies of research: physical models and ecological approach. I did not succeed in understanding well the reasons of both, and I am perplexed about their unconditional fitting with the program.

To the physical model strategy I could object that the hope of finding variations in the physical behaviour of sound sources that have a meaning in the behavioural world of the listener is open to chance.

1. Taken for granted that auditory perception is serviceable to adaptation to the environment, one has to notice that physical models cannot be found in natural environment: perfect cubes, spheres, plates do not exist in it.

2. Stimuli coming from physical sources vary along dimensions like intensity, spectral composition, duration of parts, time intervals among parts of the stimulus, profile of the envelope and so on. Percepts vary along related dimensions, like loudness, timbre, duration, articulation, dynamics and so on. But percepts vary also along other dimensions, like volume, presence, brilliance and so on – not to mention expressive contents – that have no direct conterpart in the physical stimulus.

Considering what is known in the visual domain, it is almost sure that perceptual continua are not congruent with physical continua: the perceptual dimensions of colours are other than the simple variations of wavelength. Giordano (see chapter 5), showed that discrimination between steel and glass materials is strongly influenced by the size of the objects. Small steel objects are recognized as being made of glass; large glass objects are recognized as being made of steel.

The investigation based on physical models is yet to be continued, because it brings to light unexpected effects. For instance, some manipulations of physical parameters of a unique acoustical signal of a falling body lead to the representation (a) of something like a faint explosion; (b) of the falling body and of the object on which the body is falling; (c) of the falling body, of the plate on which it falls and of the plane that bears the plate on which that body fell.

I consider this phenomenon the best result of the work of the unit of Udine, since it literally reproduces a phenomenon well known in vision, that of "phenomenal scission" or of "double representation". The paradigmatic case is that of "perceptual transparency" [132, 174], where a surface is perceived as the superposition of two surfaces, the upper one being transparent: see Figure 2.2.

Given that the perceptual fact is "a black bar on a white cross… the black bar is transparent", the phenomenal scission concerns the dark grey surface: it represents at the same time the portion of the white cross seen through the transparent bar and the portion of the black bar superposed to the white cross. Notice that on the stimulus display (distal stimulus) there is no white cross (there are 2 irregular white surfaces), there is no black bar (there are 2 irregular black surfaces) and there is an object that we do not mention: the irregular dark grey surface.

The effect is linked to the relations among the reflectances of the involved surfaces (at least four) and some topological and figural characteristics. Other instances of double representation are: scission of homocromatic surfaces, totalization, veiling, anomalous surfaces, amodal completion and figure/ground phenomenon. The common feature of all these phenomena is that they assure the simultaneous perception of objects that mutually occlude themselves when observed from a still point of view [246].

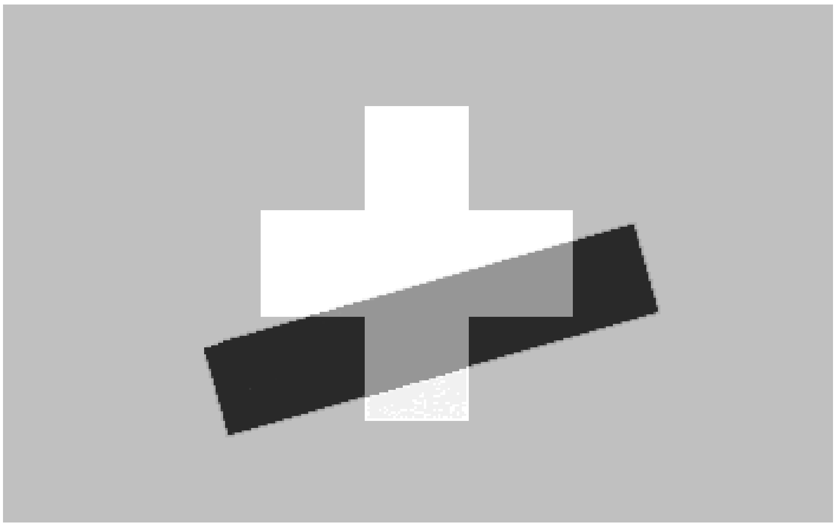


Figure 2.2: “Perceptual transparency”.

In the auditory domain we face the same problem: in the physical environment there are many objects simultaneously vibrating, so that the sound wave coming to the ear is unique. At this point there is the necessity of extracting from it the signals or the cues that can restore the multiplicity of physical objects, because choices in behaviour depend on a reliable map of objects, distances, obstacles, routes and so on. Take for instance the impact sounds: we perceive at the same time the stroke and the stroke object (a falling ball, the plate on which it falls). Now, which part of the proximal stimulus (the acoustic wave at the eardrum) goes to build the striker, and which part goes to build the striken object? In my opinion, this is a high priority goal for the research in the field.

2.7 On the ecological approach

Compared with that of physical models, the ecological approach is surely more sensible. Let us disregard all the possible physical sources of sound and noises (for the most referring to objects inexistent in the natural setting, like cubes, spheres or thin plates squared or round in form), and let us concentrate on physical objects actually present in the human environment (stones, trees, running water and so on). After all, the auditory system went to its present structure in order to make human behaviour adaptable to that environment. Nevertheless, even the ecological approach presents some problems.

I see the difficulty of the ecological approach in the neglected distinction between physical environment and behavioural environment. The absence of this distinction is for me an enigma. Gibson [103], which ecological psychologists unceasingly refer to, was a pupil of Koffka, who recognized and imposed the necessity of taking into account, when explaining behaviour, the things as they appear, and not as they are (an attitude entirely shared with ethologists, see Uexküll [247], or Lorenz [158]). Besides, Gibson [101] was the discoverer of the “pictorial perception”, that is the most flagrant contradiction with the principles of ecological perception: we see things that are not. It is true that, in the study of sounds, we have to consider sounds that take place in human environment, but this environment is not the physical environment, but the behavioural one. We react to auditory signals as their sources appear, not as their sources are.

In detail, the ecological perspective does not explain errors and illusions and consequent unsuitable behaviours (let us speak of perception as a kind of "behaviour"). As we pointed out before, we hear what does not exist, we do not hear what exists, we hear sources inexistent in the natural environment, et cetera. About errors, illusions and consequent unsuitable behaviours are an unsurmountable obstacle for those that trust in "direct" perception – there is to say that perceptual systems developed in a frame of costs and benefits. We cannot "hear" sounds below 20 Hz, and in this way we are "deaf" to the vibrations produced by physical objects whose size is rather large; at the same time, we cannot hear sounds beyond 20 KHz, and we are therefore deaf to vibrations produced by objects that are rather small. In principle, we could be supplied by an auditory system that can detect signals even below 20 Hz and beyond 20 kHz, but the benefits of this supply (avoiding perils, gain opportunities) could be achieved by unbearable costs (in terms of physiological structures).

There is another fact to point out, concerning the "ecological" adaptation of the auditory system to the physical environment. Undoubtely, there is such adaptation, but the auditory system is a procedure for processing acoustic information that has its roots in the deepest paleontological eres: they even say that cochlea is the transformation of the swimming bladder of fishes. Auditory system developed through the eres, mirroring more and more precisely the physical environment. At this point, the question is: what sort of physical environment is represented by the today human auditory system? According to ethologists, we come to light with a standard toolkit of processing abilities that refer to ecological conditions lost in distant ages, and only after the birth the auditory system specializes in interpreting signals coming from a given environment (forest noises for Amazonians, traffic noises for us). That means that we shall understand better the relation between acoustic stimuli and percepts if we shall investigate gross properties of physical environment that were important also in the past ages. Limiting ourselves to impact sounds, I expect the best results from investigating the sounds produced by concavities (holes) or convexities (bulges out), by hardness or softness of terrain, by rapid or slow running of water – not to mention the sound produced by a source coming nearer or moving away.

Coming back to direct perception so dear to Gibsoneans I ask myself in what sense we can define "direct" the perception of events (in the visual domain the movements, everything in the auditory one, the temporal dimension being unavoidable), while we often have to wait for the end of the signal in order to make use of its preceding part (the case of the discharge of the machine gun is rather simple, but there are also cases by which the succession of auditory events is different from the sequence of acoustic stimuli, see Vicario [246]. It is plain that such a phenomenon requires a sensory storage, a categorization of the signal, a recovery of previous experiences and so on, making direct experience entirely other than "direct". The only way to avoid such a criticism is to suppose that perception of events is "tuned" on physical becoming, that is on a philosophical construct, not on a fact. With such premises, if Gibson's "optical array" can be reduced to the *eidola* of early atomism (V-IV century b.C.), recent calls for the “tuning” of behaviour on physical becoming are just a form of occasionalism (XVI-XVII century a.C.).

2.8 On phenomenology of sounds

“Phenomenology” is a philosophical term [236] that refers to the analysis of the *Erlebnis* (direct experience) by means of nothing but its own contents. The classification of sounds according to their perceptual characteristics (high/low, loud/soft, brilliant/dull, threatening/reassuring and so on) is phenomenology. To classify sounds according to their spectral composition, or to the mechanical sources that produced them, or to evoked potentials in the ear or in the brain, is no phenomenology at all. As to psychology, the phenomenological attitude is that of Gestalt oriented students, who place the detailed description of percepts (and of all conscious experiences) before the explanations in terms of stimuli, neural correlates, past experience, attention, motivations and so on [170]. In the last decades took place also a method of investigation called *experimental phenomenology*, which we have to be warned about: as I demonstrated [245], experimental phenomenology is nothing but the usual scientific method, this time applied to mental contents.

As far as I know, we lack a well founded phenomenology of sounds, in a way comparable with that of colours [134]. Let us for a moment overlook the most obvious tripartition by noises, tones and speech. If we direct our attention to noises, we find only a long list of nouns (buzzing, squeaking, trampling, blowing, blasting *et cetera*, not to mention the numberless onomatopoëses), without any effort to distinguish the phenomenal dimensions of the noises themselves. It is likely that some nouns refer to different degrees of the same sensation: for instance, whispering and muttering, and that some nouns refer to the same degree of different sensations: for instance, the quiet flowing of the water of a streamlet, or the quiet skimming through a book. I found the sole efforts to create a phenomenology in manuals for police officers or jurists, who need to decide what sorts of noises (and at what loudness level) have to be considered disturbing or harmful.

The only novelty that I found in the field, is that of Fernström and coworkers (see chapter 13), that seems to me sensible and productive. As it is well known, they try to cluster various kinds of noises in a twodimensional space whose coordinates can be changed at will. In a way, the method is a refinement of the “semantic differential” [190], widely adopted in psychology to measure the meaning of any sort of mental contents. I rely on Fernström’s method to achieve a phenomenology of sounds based on their perceptual characteristics, and not on physical or physiological prejudices.

2.9 On expressive sounds

Phenomenology of sounds may appear as a huge enterprise, considering the fact that percepts have, besides the characteristics of their own (loud/soft, brilliant/dull et cetera), information about the nature of the source and its probable future behaviour. When we speak of “expressiveness” we refer to the fact that objects and events share not only *primary* qualities (size, form, weight and others, independent from the subject) and *secondary* ones (colour, smell, taste and others, rather dependent on the subject) but also *tertiary* qualities (good/bad, threatening/attracting, gloomy/happy and others) not dependent, as a rule, on the subject.

Tertiary qualities are often defined as *physiognomic*, that is "informing on the nature" of the object or of the event. In a sense, they are added to perceptual features: for instance, when listening to a sound we can say: "this is a violin", and we can also add: "this is a good violin". Indeed, objects and events present a special sort of expressive features that are *Aufforderungscharaktere* (a term translated by Gibson as *affordances*), that is information about their manageability; for instance, a door handle has a form that asks for grasping it. In general, events are far more expressive than objects: the way a man moves his hands when speaking carries a lot of information (it is defined as "non verbal communication"); the way a speaker modulates dynamics and phases of his utterances is another important communicative factor (defined as "supersegmental").

There is evidence that expressiveness in perception is not a matter of computer like processing of signals, covert thinking or previous experiences. Expressive features of objects and events are immediately manifest as form, colour, size and so on. Gestalt theorists stress this fact, producing numberless instances, at least for the visual domain.

The role of expressive features seems that of guiding the behaviour in an environment where motor responses have to be performed in suitable way and in the least time: In the visual domain we perceive that a beam is threatening to fall, or that the movements of the animal in front of us presages an aggression. From the interactive behaviour of two persons we immediately realize which is the dominant and which the obedient one. Expressive features are the foundation of social behaviour.

In the visual domain, movements are far more expressive than objects, since to the spatial features (form, colour, size) the evolution of these features in time is added. With even stronger reason, that seems to hold in the auditory domain, where there is no stimulus unless it goes on for a suitable time. As a consequence, research on expressiveness in the auditory domain should rather concern the profile of variations than the spectral composition of sounds. A lengthy experience in the perception of expressions in movements starting from the study of Heider and Simmel [119], and the theoretical and experimental contributions of Michotte [176] provides a lot of recipes for the perception of various mechanical or intentional contents. Rightly famous is the method of Johansson [126] for the recognition of locomotions.

Having worked in the field, I ask myself whether we should continue in finding stimulus conditions for all the sorts of expressive contents, or make an effort to establish the rules underlying all those phenomena. For instance, I remember the gross taxonomy made by Kanizsa and myself [133] concerning movements, and I think that it could be transposed to sounds. So we should have: (A) a first class of "natural" sounds, that is those produced by physical processes; (B) a second class of "mechanical" sounds, for the most repetitive or cyclic patterns of noise; (C) a third class of "intentional" sounds, that are those produced by living beings during their motion in the environment or during their interaction with other living beings.

Perhaps there is a simpler way for attacking the problem of expressive sounds. A careful exam of hundreds of "sounds effects", devised for adding the soundtrack to movies, persuaded me that the first and simplest categorization of sound and noises is that of “animate” versus “inanimate”. For instance, cyclic noises, or noises that vary in regular manner are perceived as coming from “natural” sources or mechanical devices; while irregular, unsteady or aperiodic noises are easily attributed to the presence of a living being acting in the environment.

2.10 Some findings and suggestions

The preceding points have been the theoretical framework for the research carried on by the unit of Udine in the “Sounding Object” project. From that framework, there emerged some findings and suggestions that I briefly summarize.

The awareness that phenomenal scission of the acoustical stimulus is the central point for understanding the multiplicity of sounding objects on the auditory scene, enabled us to imagine and to perform the experiments reported by Burro, Giordano and Grassi in this book.

The analysis of results of the aforesaid experiments makes me foresee that sooner or later we shall face the two major problems of auditory perception: (a) the formation of auditory events in the flow of phenomenal time and (b) the non deterministic and non computable relations among features of the already formed auditory events (on this point, see Vicario [242]). I hope that the knowledge of the identical matter of fact in visual field, as to the perception of objects, movements and changes, will assist us in seeing more clearly the problems of auditory perception.

The expressive value of sounds and noises came immediately on the foreground. Observations performed on noises produced by dragging objects on sandpaper, as well as produced by drills under stress, showed that expressiveness does not appear in steady sounds, and that even the recognition of their source is impaired.

The filling up of vessels by dropping water seems to give rise to two frequencies shifted in opposite direction, presumably that of the growing water column and that of diminishing air column above the surface of the water. The perception that the vessel is filling up, or that it is near to be full to the brim, depends on the slope of shifting fundamental frequency. I argue that before investigating expressiveness, it should be ascertained the conditions of the perception of change. We performed a couple of formal experiments on the perception of acceleration and slowing down of rhythms. We found confirmation of an already suspected fact [244]: the threshold for the perception of an acceleration is lower than the threshold for slowing down. The fact could be interpreted as higher sensitivity for approaching sources (incoming opportunity or peril) than for those going away (less attainable opportunities or ceasing peril).

Chapter 3

Sound objects and human-computer interaction design

Mikael Fernström
University of Limerick – Interaction Design Centre
Limerick, Ireland
`mikael.fernstrom@ul.ie`

In this chapter we will look at some ways that sound can be used in human-computer interaction and how sound objects can be designed and applied as *auditory icons* to improve usability and also offer alternatives to visual representation in user interfaces. For the interaction designer sound objects offer many new possibilities that are substantially different to using sampled sounds. We outline a roadmap towards improved design of auditory enhanced user interfaces with sound objects.

3.1 Background

Very little of our natural ability to hear is currently used in the metaphorical model worlds of human-computer interfaces, despite almost three decades of research on computer sound. Auditory interfaces can free up the visual modality, for example when using a wearable computer or mobile device [195]. Using sound is also a way to draw attention to events and to support peripheral awareness [94]; to give users confirmation that their actions have succeeded or failed [37]; to monitor processes [73]; and for notification [93].

There are many different examples of auditory interfaces (see for example [99] for a general overview). Some researchers have focused on different components of auditory interfaces, while others have worked on different domains or applications. Components are for example *earcons* [20, 38] and *auditory icons* [43, 93]. Domains can be, for example computer games sound effects, auditory display of multivariate data, auditory alerts and alarms. The components, or *widgets* as they are often called, can be regarded as building blocks for interface design, i.e. representations that we perceive the state of and in some cases can manipulate, e.g. soft buttons (graphic buttons on a display), scroll bars, check boxes, pulldown menus, icons. In this chapter, we focus on auditory widgets, in particular the two widget types *earcons* and *auditory icons*. *Earcons* are short musical motifs (sequences) and the meaning of these motifs has to be learnt. They can be organised into families of musical messages to represent for example hierarchical information. *Auditory icons* mimic or represent everyday sounds that we might be familiar with from our everyday experience in the real world, hence the meaning of the sounds seldom has to be learnt as they metaphorically draw upon our previous experiences with the real world. A fundamental difference is that *Earcons* are abstract representations while *auditory icons* are analogical representations. They can, of course, be used in combination in hybrid interfaces [6]. Gaver [93] pioneered the field of auditory widgets with his *SonicFinder*. It extended Apple's file management application *Finder* using *auditory icons* with some parametric control. The strength of the *SonicFinder* was that it reinforced the desktop metaphor, creating an illusion that the components of the system were tangible objects that you could directly manipulate. Apple's later versions of MacOS implemented *desktop appearance settings* including sound, inspired and informed by the *SonicFinder*. Gaver's work did not address the usability of auditory interfaces. He demonstrated that it might be a possible route to explore for a richer user experience. On most personal desktop computers today, users can activate *schemes of desktop sounds*. These schemes are often hybrids between *earcons* and *auditory icons*, i.e. abstract information represented by *earcons* while auditory icons can represent information that can be regarded as metaphorically analogous to real-world events. A thematic metaphor can hold each scheme together. Existing sound schemes are currently based on playback of sound files, without any parametric control.

On wearable computers and mobile devices the exploration of auditory interfaces has only recently started (see for example [35]). While desktop computers now have ample processing power and high quality soundcards, mobile and wearable computers have less processing power and seldom-adequate sound capabilities. Most wearable computers and mobile devices have small visual displays with limited capabilities. The use of such devices is often characterised by the user having to devote shared visual attention elsewhere in their environment. Up until recently handheld computers, such as Palm Pilot$^{\mathrm{TM}}$, had extremely limited and default sound schemes. When a soft button on the touch/stylus sensitive display is activated the visual button icon is reversed in greyscale. When the button is released, a click sound is produced to indicate successful activation. Brewster [35] showed that by extending the sound schemes with three different simple earcons for the different states of soft buttons (button down, button up and button slip-off) both usability and efficiency was substantially improved. To date, the most successful applications of *earcons* (in a wide sense) are probably handheld devices such as mobile telephones. The visual displays on these devices are small and the use is substantially different to desk-based applications, including that people carry telephones in pockets and bags where the visual display cannot be seen. A few years ago, before *personalised ring tones* were introduced, every mobile phone owner within earshot reached for their phone no matter whose phone was ringing. Today, many users have personalised ring tones and different melodies for normal calls, calls from specific persons, notification about text messages, etc. Another area in which *earcons* have been successfully applied is in Computer-

Telephone Integration (CTI) applications, e.g. voice mail systems. As suggested by Leplatre and Brewster [150], the hierarchical structure of menu systems in CTI systems can be represented by *earcons*. For the novice user both *earcons* and voice prompts are played. When the user gets more familiar with the system, it is sufficient to play just the *earcons* to let the user know where their actions have taken them in a functional hierarchy, success/failure of operations, etc.

Several studies have shown that that auditory widgets can improve usability and performance [100, 94, 39, 19]. Unfortunately little of this has transpired into our contemporary computers and few users leave their sound schemes turned on. Some users claim that the sound schemes do not contribute to the quality of use and that the sounds are annoying, especially in offices with more than one person. The general noise-level increases and perhaps colleagues find each other's sounds disturbing, confusing and annoying (despite phones ringing, ambient conversations, MP3s played, air conditioning, traffic noise, etc). Another possible reason that users turn off the sound on their computers might be that the desktop metaphor interface has already been optimised for the visual modality, hence replicating representation in the auditory modality creates redundancy, which might be of some value, but more often be felt to be superfluous.

3.2 How to design more usable auditory interfaces?

To improve the design of auditory interfaces and in particular auditory widgets, a number of issues have to be addressed. First of all, we have to consider where and how auditory interfaces are appropriate. We have to take into account the users' capabilities, while carrying out tasks in real environments, and also consider that surrounding noise levels might mask the sound of auditory widgets. Where sound might enhance interaction, ways of creating and testing auditory metaphors need to be explored. Also, as suggested by Gaver [98], in many cases we need parametric control. Then, small events can make small sounds; the effort of an action can be dynamically represented, etc. Perhaps one of the problems with the sounds used in auditory interfaces is that they always sound the same, all the time. As Barry Truax points out [235], before the development of sound equipment in the 20th century, nobody had ever heard exactly the same sound twice. He also remarked that fixed waveforms, as used in simple synthesis algorithms, sound very unnatural, lifeless and annoying.

Just as visual graphical user interface design benefits from the skills of graphic designers, we need to consider the art of sound design, as suggested by Eric Somers for auditory interfaces [226]. Sound designers, also known as *Foley artists,* in radio, television and cinema production have shown that well-designed sounds can enhance our experience. It is also a well known fact among professionals involved in sound design and the recording industry that some sounds never sound real when recorded, e.g. the recorded sound of a gunshot often sounds like a popping cork. To get a sound that really sounds like a gunshot, a sound designer has to create layer upon layer of different sounds that meld together *sounding like* the real thing, in playback. The reasons behind this are highly complex and go beyond the issues in this chapter, but, by studying the art of *Foley*, some critical design problems might be resolved, e.g. the shortcuts and *cartoonification* that a sound designer might use when creating a sound that really *sounds-like* rather than modeling the real complex physical event.

3.3 The novelty of sound objects in interaction design

Sound objects can be used for parametrically controlled *auditory icons*. For interaction designers, the novelty of having sound objects is that suddenly the auditory domain can become as live and animated in direct manipulation interfaces as the visual interfaces are. To understand what sound objects can do for interaction design we need to understand what is different with sound objects compared to traditional playback of sound files or simple sound synthesis. In an unfinished book manuscript, Buxton, Gaver and Bly suggest a useful categorisation [44]. They distinguish between *fully formed objects* and *evolutionary objects*. With the former category, all variables are known at instantiation, hence when an object is activated a sound is produced, from start to end. With the latter category, variables controlling properties of the sound are updated while the sound is playing. To further clarify the difference between these two concepts you might consider an analogy from the musical world: Striking a key on a piano, after which you have no more control over the sound, compared to playing a note on a violin where you have full expressive control over pitch, loudness, timbre, vibrato, etc., throughout the life of the note.

3.3.1 Direct manipulation revisited

With sound objects, we can create both evolutionary and fully formed objects, but it is the evolutionary aspects that make sound objects particularly interesting. The interaction style of direct manipulation has proven to be extremely successful over the past two decades. It has offered users systems that are easy to learn and master, giving more confidence and enjoyment of use [217, 124]. So far, research on direct manipulation has mostly focused on visual representation of objects. Some of the key features of direct manipulation are:

- Visibility of the objects of interest.
- Continuous representation of objects.
- Physical actions on objects instead of complex syntax.
- Rapid and incremental reversible operations.

When designing interaction with both visual and auditory representations we also have to take into account that vision and audition have different characteristics. Some information may be better represented with sound, while using both vision and audition may be better in other cases. Gaver [93] suggests a simple contrast between vision and audition; we see spaces and objects and we hear events. Combining the key features of direct manipulation with this implies that sound objects might lend themselves best to represent actions. In previous work on auditory interfaces, human action has to a large extent been treated in a discrete way, e.g. like kicking a football, where a user action starts a process that then completes without user control. An alternative view is actions as a continuous flow, e.g. a pen stroke, where we continuously move a pencil on a surface, both relying on our learnt gesture through proprioception, as well as haptic, visual and auditory feedback.

3.4 A roadmap to interaction design with sound objects

If and when auditory widgets are considered to be beneficial for a particular interaction design, a number of investigations and design decisions lie ahead. We need to identify sounds that might contribute to the users' performance and subjective quality of use. It is also necessary to consider how we create and test auditory metaphors. In the following section we outline a number of steps for developing auditory enhanced widgets and we describe some practical cases.

3.4.1 Choosing sounds and metaphors

One often used method to find sounds that potentially can be used in user interface design is to conduct listening tests with recorded sounds, testing how participants can identify what they hear. This is also useful for increasing our understanding of what people think they hear when exposed to isolated auditory stimuli, with or without any particular context [14, 93]. Another approach was used by Barrass [16] who created a case-based tool based on nearly 200 anecdotes, collected via email, about situations where sound played an important role in real-world situations. In our own work we have used the listening test approach. We collected 106 recordings of everyday sounds, based on ideas from Gaver's [96] trichotomy of liquid-solid-gas sound classification. We had 14 students to listen to the sounds with headphones and respond by writing down a sentence for each sound describing what they heard. We also asked them to state how familiar they were with each sound. Our results correspond quite well with other studies such as [13] and [109]. The results indicate what sounds that are easy to identify and classify. In the case of all users giving an almost identical response, we can be fairly sure that the sound in question is easy to identify. The actual texts that our participants provided can act as a free-form description of what the sounds can be used for, either directly or in the design of an interface metaphor.

In the following sections, we look at some practical design examples.

Progress or upload/download indicator

The sound of containers, such as bottles or glasses, being filled with or emptied of liquid is a robust sound in terms of identifiability of both action and agent. Several studies have ranked water sounds as highly recognisable [13, 109]. This also corresponds to our own investigations. From our collection of 106 recorded sounds, of which 16 were various water sounds, the accuracy in identification of filling and emptying liquid was 93%.

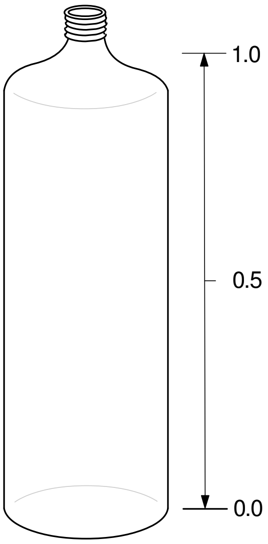


Figure 3.1: Normalisation of fill/empty levels.

We then designed an exploratory investigation to further investigate this particular type of sound. 197 undergraduate computer science students participated in this study. The sounds used in the experiment were 11 digitally recorded sounds of water being filled or emptied, at different rates, using 0.5 and 1 litre plastic bottles. The participants used headphones for listening to the sounds. They were instructed to respond using their keyboards. The experiment was divided into three sections: to detect filling or emptying; to detect half full/empty; and to detect almost completely full/empty. The order between sections was randomised, as was the order between sounds in each section of the experiment. When we asked the participants to respond if the sound was filling or emptying, 91.8 percent responded correctly for emptying sounds and 76.4 percent for filling sounds. In the section where they responded to when the bottle was half full or empty, responses during filling had a mean of 0.40 (range normalized to 1, see Figure 3.1) with a standard deviation of 0.13. During emptying, responses had a mean of 0.59 with a standard deviation of 0.18. In the section where users responded to whether the bottle was almost full (just about to overflow) or almost empty, the mean value for filling sounds was 0.68 with a standard deviation of 0.18, and for emptying sounds the mean was 0.78 with a standard deviation of 0.18. These results indicate that the effect is quite robust and can potentially be used, for example, as an auditory representation of progress, upload/download, or perhaps even for scrolling. Based on our findings, we analysed the sounds and modeled the sounds as a sound object with various degrees of *cartoonification* [97], i.e. to simplify the model and exaggerate selected features of the sounds.

Soft buttons with both visual and auditory representation

A common problem with soft buttons in graphical user interfaces is that users sometimes slip-off a button before releasing it, which normally results in the function that the button is expected to activate does not get executed. This feature is sometimes useful, however, as it allows the user to consider if a function should be activated or not, i.e. if the user changes his or her mind with a soft button pressed, the button can be released by slipping-off, to avoid execution of the button's function. In terms of the key features of direct manipulation, this accommodates reversibility of action to a certain degree. As graphical interfaces get smaller, on handhelds and wearable computers, the problem of accidental slip-off increases. Brewster [36] has shown that having different sounds for button-down, button-up and slip-off on a Palm III Pilot$^{\text{TM}}$ handheld computer substantially enhances usability. Our own informal tests fully supports Brewster's findings, but it is interesting to note that by using sound objects, we only need one single sound model to produce several different kinds of click sounds. As a contrast between the use of sampled sounds and sound objects, consider the following: With sampled sounds, each sound on average 400 milliseconds long and sampled at 44,100 samples per second in 16-bit mono, each sound file will be at least 34 KB, hence three different click sounds requires 102 KB memory space. Our sound object model for impact sounds only requires 56 KB, and can produce almost any kind of click sound depending on the parameters passed to the model. As memory space is scarcer on handheld computers than desktops, the sound object approach is more efficient and effective.

Click sounds, or rather sound object impact model sounds, can also be useful for other kinds of widgets. An impact model can be used in a maracas-like way, i.e. when the user shakes the device a number of impact sounds proportional to the amount of battery remaining can be produced.

Soft buttons with only auditory representation

If we cannot see an interactive soft widget, how would we represent it? There are several interesting developments in interface design for visually impaired users. Some of these developments are also useful for interfaces where the visual modality is needed for other tasks. Buxton [42] and Boyd et al. [26] outlined some of the issues when designing for impaired users, especially since graphical user interfaces (GUI) have become so dominant. Other researchers have tried to address this problem by trying to represent GUIs completely with sound, e.g. Mynatt's *Meractor* system [183] using auditory icons, synthetic speech, spatialisation and abstraction of widgets on a semantic level rather than a direct remapping from the visual interface. As direct manipulation is central to interactive GUIs, Winberg and Hellström [255] addressed this in an auditory version of the game *Towers of Hanoi*. They used distinct timbres and stereo panning to represent the objects (discs) and their location. The objects were constantly represented, i.e. sounding all the time. Based on their findings, Targett and Fernström [232] created auditory versions of two other popular games, *X's and O's*, and *MasterMind*, using earcons and auditory icons to represent object identity, location and all the complex states of each game. Sounds were only produced in conjunction with user actions. The games were tested and evaluated with five users who all reported the games to be fun and engaging. Despite the fact that representation of objects was only transient, in conjunction

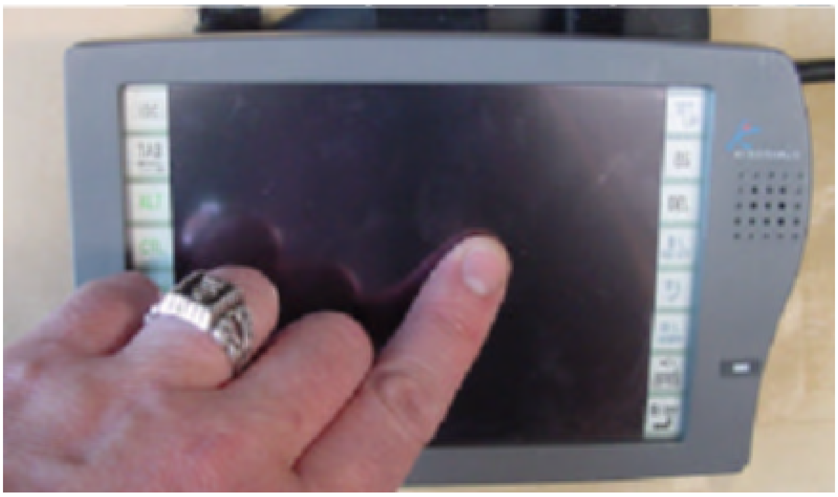


Figure 3.2: Xybernaut touch device.

with user actions, the users didn't have any problem picking up the state and location of the objects.

To further explore these possibilities, in particular for wearable or handheld computers, a simple prototype was implemented on a Xybernaut MA IV wearable computer. This kind of wearable normally has a system unit and battery fitted to a belt around the user's waist and various peripherals connected by cables and fitted on the user's arms and head. Some examples of peripherals are headmounted displays, headphones or headset, LCD display with touchinput and membrane keyboard to be fixed on the arms of the user. Typical scenarios of use, suggested by Xybernaut, is aircraft and telecom maintenance, courier service and mobile customer assistance work.

In our prototype we used the touch sensitive display subsystem (see Figure 3.2), which is normally worn on the arm, fixed to the user's belt around the waist. The size of the touch area is 120 by 90 millimetres. We only used the touch detection of the device, not the visual display. A number of soft button areas, with different layouts, were defined. To represent that the user is moving fingers on a button area a simple friction sound model was used, based on a noise source and band pass filters. To emphasise the boundaries of each button, an impact model was used for producing entry and exit click sounds. See Table 3.1 for the complete mapping between actions and sounds.

In an informal test with three users we found that they were able to feel their way around the touch device and make drawings of the layout of soft buttons, see Figure 3.3. The sounds were heard in mono, using a simple headphone in one ear. Configurations with three to six buttons in various configurations were explored. This indicates that this kind of auditory representation of soft buttons allows users to have a pseudohaptic experience giving them a mental model of the layout of the device. Further work is needed to refine and validate this type of widget, but it has the potential to be both useful and versatile. One can, for example, add other auditory icons to represent the actual functionality of each button, in addition to the sounds representing the buttons themselves. This type of widget frees up the visual modality, hence it might be particularly useful

Action	Sound	Function
No touch	No sound	
Touch area outside button	No sound	
Enter button area	*Click* sound	
Move finger on button	Friction sound	
Exit button area	*Clack* sound	
Lift finger off button	*Tock* sound	Select/Activate function

Table 3.1: Sounds for the prototype implemented on Xybernaut. The *italics* indicate onomatopoeic description of the sounds used.

in wearable applications where users need vision to physically navigate the environment.

Using sound to represent handwriting gestures

Another application area for friction sounds is for improving handwriting recognition systems such as Palm's *Graffiti*, which allows the user to make simple, handwriting-like gestures with a stylus on a touch sensitive area of a Palm Pilot handheld computer. If the user's gesture has been reasonably correct, a corresponding character is added to the current cursor position on the screen. Most Palm users find that when their *Graffiti* writing speed improves, they eventually find that perhaps one in every twenty characters is misinterpreted by the system. This could be due to a lack of feedback while making the gesture with the stylus. If the user was given continuous feedback while making a handwriting gesture with the stylus, this problem might be reduced. To attempt to remedy this, one can add a friction sound while the user making a *Graffiti* gesture, to provide continuous feedback about the user's action.

3.5 Implementation issues

In our prototypes described in this chapter, Macromedia Flash was used for programming interaction sequences and Pure Data's `pd` engine was used for modelling and making the sounds. Control parameters were passed from the Flash applications to `pd` patches containing the sound models. While this is a fast and convenient way to do rapid prototyping, a fully embedded system probably requires to be more closely integrated with the operating system.

3.6 Summary

In this chapter we have shown how sound objects can be applied to enhance usability, both as a complement to visual user interfaces and as an alternative to visual representation. Sound objects offer a more versatile and economical way to create interaction sounds, compared to sampled

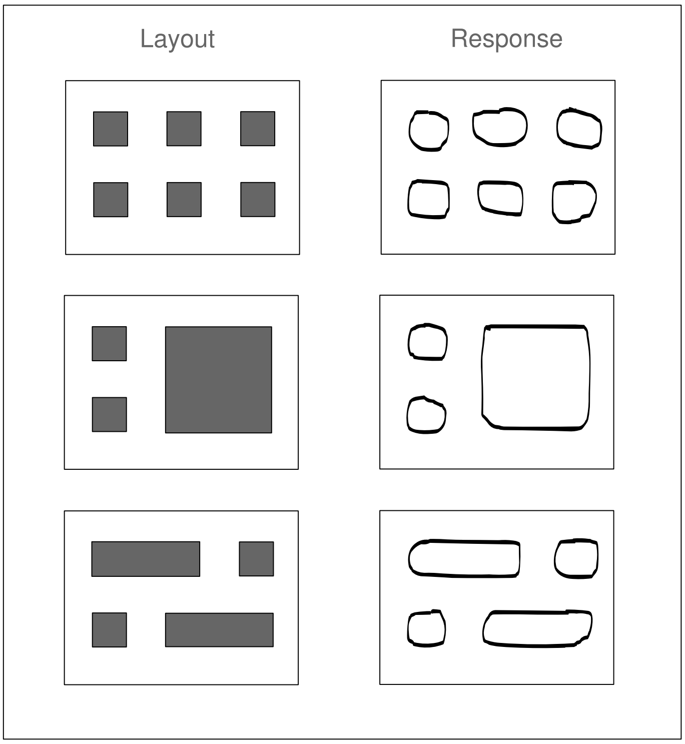


Figure 3.3: Examples of soft button layouts and responses.

sounds. For example, with the sound object impact model almost any impact sound can be produced, ranging from wood to rubber to metal to glass just by changing the parameters controlling the model. We have also considered how sound objects are substantially different to sampled sounds as they offer full parametric control, as evolutionary objects, throughout the duration of a sound. A number of methods have been described for the selection and evaluation of sounds for application in human-computer interaction design.

3.7 Acknowledgments

Thanks to Liam Bannon, Eoin Brazil, Luigina Ciolfi and Sten-Olof Hellström for their reviews and creative comments on this chapter. Thanks to Xybernaut GmbH for giving us access to their wearable computers.

Chapter 4

Impact sounds

Massimo Grassi and Roberto Burro
Università di Udine – Faculty of Education
Udine, Italy
Università di Padova – Department of General Psychology
Padova, Italy
grassi@psy.unipd.it, burro@unipd.it

4.1 Introduction

Recent researches demonstrated that listeners are able to extract from a sound the physical properties of the objects that generated it. The results of the experiments are interesting for two reasons. Firstly, the sound reaching the ear is the result of the contact between, at least, two objects. Nonetheless, listeners are able to segregate from the sound the information concerning the sole object they are asked to evaluate in the experiment. Moreover, after listening to the sound, listeners provide veridical estimations of the static properties of the objects that generated it. So far, listeners were able to estimate the length of a rod dropped on the floor [49], the ratio dimension and the shape of either struck bars or plates [145, 142]. Furthermore, participants scaled correctly the hardness of a set of mallets striking cooking pans [82], and the gender of walkers from the sound of their footsteps.

In the experiments, participants often describe directly the sound source event (the distal stimulus) rather than describing the acoustic parameters of the sound that the source is producing (the proximal stimulus) [92, 240, 239]. Moreover, the best predictors of the performances are the physical properties of the distal stimulus (length, shape, material, etc.) rather than the acoustical indexes of the proximal stimulus (for example, sound pressure level, frequency content, etc.). According to Gaver [97, 95] and Fowler [79, 80] we perceive sound sources because the source properties have consequences for our behavior. In addition, Gaver [97, 95] purposed the distinction between everyday listening and musical listening. Everyday listening corresponds to the direct perception and the recognition (either veridical or not) of the sound source event: the distal stimulus. On the contrary, musical listening does not involve a recognition of a sound source event but rather a simple listening to the properties of the proximal stimulus: the acoustic wave.

However, the relationship between physical interactions, sounds and human perception needs to be further analysed. For example, impact sounds are the result of the interaction between, at least, two objects. The resulting sound is dependent on which of the objects is vibrating. Impact sounds can be produced in a number of ways and, in any specific impact, either one object or the other (or both) is put into vibration. If we listen to the sound of a book dropped on a concrete floor the air pressure wave is mainly produced by the vibration of the book. On the contrary, the same book dropped on a table will produce a sound that is the result of the vibration of the book and the vibration of the table. Also the strength of the impact can affect significantly the resulting sound: if we drop the book on the floor from a height of twenty centimetres or forty centimetres the resulting acoustic waves will be different. A higher strength in the impact will result in a different solicitation of the so-called modes of vibration of the object [180] and, consequently, in a different acoustic pattern. In fact, when put into vibration each object vibrates in a limited number of ways, specific of that object. The modes of vibration are the possible patterns of vibration of a given object. Consequently, the sound produced by any object is peculiar and is characterised by only certain frequencies and, therefore, by a particular timbre.

In the researches reported above, the distinction between the two partners involved in the impact and their roles on the resulting sound was only marginally highlighted. Likely, the more one object contributes to the resulting sound, the more information gives to the listener about itself. Therefore, if we drop a book on the floor or upon a table we will face two quite different sounds: in the first mainly the book is vibrating; in the second both the book and the table are vibrating. As a consequence, the first sound will carry much information about the static properties of the book itself (mass, size, material, etc.) and not as much about the floor (its area, its thickness etc.). Conversely, the second sound will provide us a number of information about the physical characteristics of both the book and the table. All researches performed so far were oriented on investigating the perception of the most vibrating object: a rod dropped on the floor [49], a plate struck by a pendulum [142], a struck bar [145].

In physical models the role played by each object within the impact is identified with names. The object that provides the energy for the vibration of the second object is called exciter. Instead, the object that receives the energy from the exciter is called sounding object (or resonator). The two names indicate which action is due to each specific object: the former put into vibration the latter. So far, we saw that researches has been concentrated on investigating the perception of the physical properties of the sounding object [49, 142, 145]. Freed [82] performed the only research investigating the exciter and his research was focused on the perceived hardness of a mallet striking a cooking pan. This chapter wants to provide further evidences on the perception of exciter's properties in impact sounds. All experiments reported show studies about the perception of the less vibrating object within the impact. The experiments performed investigations where the exciter (either a wooden or a steel ball) is dropped upon the resonator (either a backed clay plate or a wooden board). The goal of the experiments is to investigate how listeners extract physical properties of the less sounding object.

In the following section we will describe the physical event: the interaction between a ball dropped upon a flat surface. Furthermore, we will describe the possible consequences of the impact on the resulting sound. The analysis of the physical interaction between the ball and the plate and its effect on the resulting sound will provide a framework for understanding how listeners perceive from sounds what Gibson calls *useful dimensions of sensitivity* (i.e. size, weight and distance) [102].

4.2 Analysis of the physical event

The event is characterised by a ball of mass m falling on a plate (or a wooden board) from height h, subject to gravity acceleration g. When the ball touches the surface all the potential energy

$$E = mhg \tag{4.1}$$

is converted into kinetic energy. After the impact, part of the potential energy is transformed into acoustic energy, part is reconverted into kinetic energy (the ball bounces, the surface vibrates) and part is dissipated. In the following two sections we will describe the effect of such interaction on the resulting sound.

4.2.1 The effect on the resulting sound

The mass of the ball (m) and the height from which the ball is dropped (h) are directly related to the potential energy of the physical impact. Therefore, since the energy is the capacity for doing work, power is the rate of doing work over time, and work, in the event under analysis, is the displacement of the air that is causing the sound, a monotonic variation of the potential energy (E) will mainly affect the power of the sound reaching the ear: the higher the potential energy of the impact the higher the power of the sound. Moreover, further differences in sounds produced by interactions with different potential energies are theoretically predictable. Firstly: the greater the energy of the impacting ball the greater the oscillation of the surface. Therefore, both the amplitude of the resulting sound and its loudness will be greater. Secondly: the greater the energy of the impacting ball the longer the duration of the oscillation of the surface and, consequently, the duration of the sound. Although this can be theoretically true, bounces succeeding the first impact can damp the vibration of the surface and therefore shorten its oscillation. Thirdly: the greater the energy of the impacting ball the longer the time of contact between the ball and the plate [8, 11]. The time of contact between the exciter and the resonator can alter the frequency content of sounds produced with monotonic variations of either m or h. A long time of contact would damp vibrations of the surface whose periods are shorter than the time of contact itself. Consequently, the higher the energy E the higher would be the damping of the high frequency modes of vibration of the surface with a resulting attenuation of the high frequency components. As a result, the sound produced by low energy interactions is bright while the sound produced by high energy interactions is dull. This difference can be recorded with the spectral centroid [106]. This acoustical index is usually correlated with the brightness of the sound.

In situations where the resonator is kept constant and only m or h are manipulated, the difference in the resulting sounds will be dependent on the interaction between exciter and resonator as described above. However, in some of the researches reported here, the resonator and the exciter are both changed within the experiment. In such situations the resonator will affect much more the resulting sound. The resonator is the object that vibrates and its vibration characterises the timbre of the sound. Therefore, any change in the properties of the resonator (i.e. material, shape, area, etc.) will affect greatly the resulting sound. In the first and second research the area of the resonator has been manipulated. The effect of this manipulation is predictable: the higher the area of the resonator the lower the frequency content of the sound. In fact, larger resonators will carry vibrations whose wavelengths are longer and, consequently, lower in frequency, than those carried by smaller resonators.

4.2.2 High order structure

The impacts investigated were partially elastic. This means that, after the first impact, the ball returns to the air and, successively, it falls again for a certain number of times. For this reason, according to the distinction proposed by Warren [251], the resulting sound can be described in two ways: from the point of view of its elementary properties and from the point of view of its *high order structure*. Elementary properties are all those properties that characterise the acoustic signal (i.e. level and spectral pattern, etc.). These properties have been already discussed in the previous section. Instead, the *high order structure* is the distribution of the bounces over time. In a given interaction, all the bounces of the ball will produce sounds very similar one another. Therefore, the *high order structure* provides us the information about the physical interaction per se and not (or not as much) about the physical properties (i.e. size, weight, etc.) of the objects involved in the event. For example, within the class of possible interactions between a ball and a flat surface, the structure of the temporal pattern permits to distinguish between a bouncing event and a rolling event (see chapter 8). Nonetheless, some difference can be found in the bouncing patterns of balls with different masses (or different height from which they are dropped) when dropped upon a surface.

In fact, the ball travels from the height h, subject to gravity acceleration g, and impacts the surface at a velocity v_i:

$$v_i = v_0 + gt \quad , \tag{4.2}$$

where t is time and v_0 is the starting velocity ($v_0 = 0$ in the event in analysis). Therefore, since time can be calculated as:

$$t = \sqrt{\frac{2h}{g}} \quad , \tag{4.3}$$

then (4.2) can be rewritten as such:

$$v_i = gt = g\sqrt{\frac{2h}{g}} = \sqrt{2gh} \quad . \tag{4.4}$$

Consequently, the impact velocity of the ball v_i is independent from its mass and increases with the square root of h. The velocity of the ball when it bounces back in the air is dependent only on its impacting velocity v_i and on the elasticity[1] e of the interaction:

$$v_{out} = -v_i e \quad , \tag{4.5}$$

where v_{out} is the velocity of the ball when it bounces back into the air. When the ball bounces back into the air its velocity will change on time as follow:

$$v_b = v_{out} + gt \quad , \tag{4.6}$$

where v_b is the velocity of the ball during the time t when the ball bounces back into the air. When the balls reaches the peak of the $1th$ bounce its velocity will be null ($v_r = 0$) therefore:

$$t_p = \frac{v_{out}}{g} \quad , \tag{4.7}$$

where t_p is the time between the impact and the moment on which the ball reaches the peak of the $1th$ bounce. When the ball reaches again the surface its final velocity will be again v_{out} and, consequently:

$$t_b = 2\left(\frac{v_{out}}{g}\right) \quad , \tag{4.8}$$

where t_b is the time between the first impact and the second impact. Hence, in impacts where h is varied, velocities v_{outs} involved are different and, consequently, also t_b will be different: the higher h the longer the time t_b. This means that bounces of balls dropped at elevated heights will be more spread in time than bounces of balls dropped at moderate heights.

Similarly, in impacts where h is kept constant and the mass of the ball m is decreased bounces will be once again more spread in time. In fact, the velocity v_i is independent from the mass of the ball m. Furthermore, the higher the mass m the longer the time of contact between the ball and the plate. As a consequence, the dissipation of kinetic energy for long time of contact will be greater than the dissipation for short time of contact. Therefore, the v_{outs} of light balls will be greater than v_{outs} of heavy balls and also the time between the first and the second impact t_b will increase. The analyses can be repeated identically for all the bounces following the first.

[1] Elasticity is a parameter that can vary from $e = 0$ (no elasticity) to $e = 1$ (perfect elasticity). This parameter is also called coefficient of restitution.

4.3 Overview of the researches

All investigations reported in this chapter will analyse the resulting sound from the point of view of the framework presented in the previous section. In particular, goal of the first experiment was to investigate the role of the *high order structure*, the amplitude and the frequency content of the sound in perceiving the size of the exciter. Goal of the second experiment was to understand whether listeners can give metrical estimations of the size of the exciter. Goal of the third experiment was to study the relationship between mass, height and distance, in the perception of the physical properties of the exciter. Methods and results of all researches will be presented in the following sections. A general discussion of all findings will follow in the final section.

4.4 The effect of sounds' manipulation

The aim of the experiment was twofold: on the one side we wanted to investigate whether the perception of exciter's size was possible. On the other side, we wanted to test the validity of the framework proposed in the previous section.

So far, results of experiments investigating the subjective size of the sounding object demonstrated that listeners are able to evaluate the size even with no foregoing information about the physical event. Furthermore, in the experiments, listeners were able to provide metrical estimations of the size of the objects [49, 142]. However, the exciter contributes only slightly to the resulting sound and, therefore, listeners face a stimulation where the sound of the object they have to evaluate is hidden by the sound of the other object. As a consequence, the perception of the size of the exciter should be, in principle, more difficult than the perception of the size of the sounding object. For this reason, we opted for an experimental procedure easier than those used by [49, 142]. Firstly, before the experiment, listeners were informed that sounds were produced by dropping four different balls onto three different plates. In addition, the task required from subjects was less demanding than those used in [49, 142]: during each trial participants were asked to evaluate which of the four balls produced the sound they were listening to.

Another aim of the current experiment was to test the efficiency of the framework proposed in the previous section. Sounds were manipulated in a number of ways. In one condition, sounds were deprived of all the bounces succeeding the first impact. Results obtained with this manipulation highlight the importance of the *high order structure* in perceiving the size of the ball. Furthermore, the role of the amplitude content of the sound (i.e. its power) and the role of the frequency content of the sound was investigated. In fact, the exciter should affect mainly the amplitude content of the sound and only slightly its frequency content. For this reason, a manipulation of the amplitude content should impair listeners' performance more than a manipulation of the frequency content.

In conclusion, the current experiment wanted to provide an explorative study on the perception of the size of the exciter. We expected an impairment in the performance for all those conditions where the sound was manipulated compared to the performance where the sound was not manipulated. Furthermore, we expected a linear decrement in the performance when the manipulation involved characteristics that are non important for the estimation of the size. On the contrary, we expected a non linear decrement in the performance in those cases where important characteristics for the estimation of the size were involved.

Method

Stimuli and apparatus Stimuli were obtained by recording the sound of solid wooden balls (pine) dropped onto baked clay plates. Balls could weight either 0.33, 1.2, 2.7 or 5.2 grams and plates had a diameter of either 160, 185 or 215 mm. Balls had a diameter of, respectively, 10, 15, 20, 25 mm. Balls were dropped manually from a wooden frame in the centre of the plate from an eight of 300 mm. Sounds were recorded with a Tascam DAP1 DAT recorder and a Sennheiser MKH 20 P48 microphone in a quiet room. Each combination of ball size and plate diameter was recorded five times. Recordings were then converted into sound files at 44.1 kHz sample rate and 16 bits resolution. One control condition and four experimental conditions were produced: (1) original sounds (control condition); (2) sounds without bounces: all the bounces after the first impact were deleted; (3) sounds equated for RMS power; (4) sounds low-pass filtered with a Butterworth filter of the $4th$ order and cutoff frequency at 5 kHz; (5) sounds high-pass filtered with a Butterworth filter of the $4th$ order and cutoff frequency at 5 kHz.

Manipulations (3), (4) and (5) were performed on sounds without bounces. After the filtering (4, 5), all sounds were equalised to their original RMS power. During the experiment sounds were played monophonically through a Yamaha sound card with a Pentium 800 computer into a Sennheiser HD 414 headphone.

Procedure Sixteen undergraduates participated individually in the experiment on voluntary basis. They all reported having normal hearing. The experiment was divided in two sessions of three conditions each. Each session began always with the control condition followed by two experimental conditions. Within the next day, subjects performed the second session where the control condition and two remaining experimental conditions were ran. The order of the experimental conditions was counter balanced across subjects. The task consisted in categorising the dimension of the ball in a four alternative forced choice task. Before the beginning of the experiment, a selection of stimuli was played to the subject. In particular, subjects heard the effect of the size of the ball on the sound quality (four balls dropped upon the same plate), and the effect of the plate diameter on the sound quality (same ball dropped onto the three plates). These examples were played until the listener was satisfied.

Results

Subject answers were converted into percent of correct categorisation. A 4 (size of the ball) × 3 (plate diameters) × 6 (experimental conditions) ANOVA was conducted on the percent of correct categorisation. A contrast between the two control conditions did not show a significant effect: subjects did not improve in the performance by repeating the control condition ($F < 1$). For this reason, results obtained in those two blocks were averaged and a new $4 \times 3 \times 5$ ANOVA was performed on the data (see Figure 4.1). The plate size did not show any effect on the

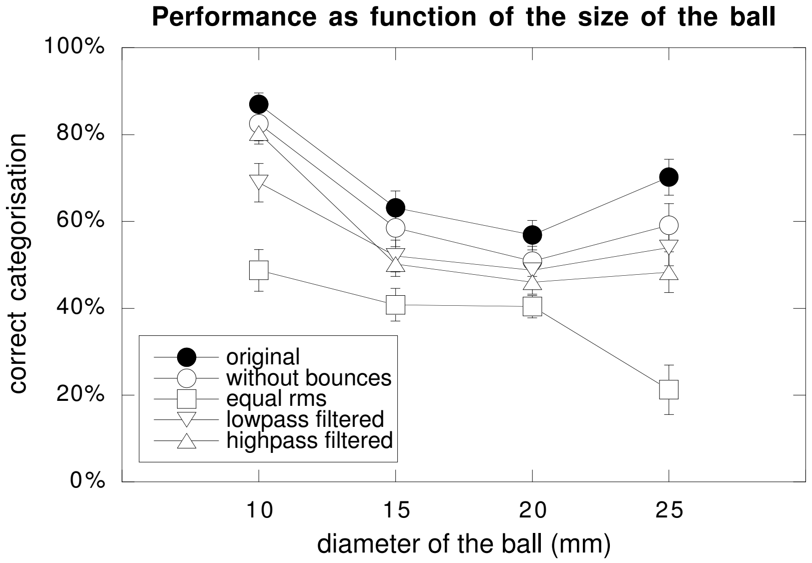


Figure 4.1: Correct categorisation of the size of the ball as a function of the size of the ball. The lines show the performance in the five experimental conditions. Vertical bars represent the standard error of the mean.

performance: subjects' performance was independent from the plate where balls were dropped $F < 1$. The performance in all experimental condition was worse than in the control condition: $F(4, 60) = 36.03,\ p < .0001$. In particular, the manipulation of the high order structure (without bounce condition) decreased by a constant proportion the performance in the categorisation: $F < 1$. Also the performance with low-pass filtered sounds and high-pass filtered sounds decreased by a constant proportion in comparison to the control condition: respectively, $F(3, 45) = 1.42,\ p > .05$ and $F(3, 45) = 2.56,\ p > .05$. However, the pattern of categorisation changed for the sounds equated for RMS power $F(3, 45) = 7.98,\ p = .0002$. In spontaneous reports after the experiment, listeners gave rich descriptions of the characteristics of the original sounds. Many of the participants noted that smaller balls produced more bounces than heavier balls. Furthermore, they reported that smaller balls produced brighter and less loud sounds.

Discussion

Listeners showed a good skill in categorising the original sound: the average performance was about 70%. This ability was independent from the plate where the ball was dropped. This result shows that listeners can evaluate the size of the less vibrating object in the physical interaction. In the without-bounce condition, the performance was slightly poorer than in the control condition and the highest within the experimental conditions. Moreover, the response pattern was similar to the pattern of the control condition. Likely, the bounces succeeding the first impact and their temporal distribution are important for the recognition of the event per se and only marginally important for evaluating the size of the ball. By listening to bounces after the first impact listeners can distinguish between a ball dropped on a plate and, for example, a ball rolling on a plate. Warren and Verbrugge [251] already demonstrated that the temporal structure is responsible for a correct identification of breaking events vs. bouncing events. In the remaining experimental conditions the performance was impaired. However the response pattern was still similar to the control condition: this manipulation was impairing the overall performance without changing the subjective size of the balls. Low-pass filter sounds are quite common in everyday life: when, for example, we listen to something through a wall the acoustic pattern is low-pass filtered. The performance with the high-pass filtered sounds was also poorer. In this condition listeners often categorised the sound as being produced by the lightest ball. In fact, the percent of correct categorisation for the lightest ball was as high as in the control condition. However, the response pattern for this condition was again similar to the response pattern in the control condition. Performance with sounds equated for RMS power was the poorest ($\sim$37%). These sounds were sounding similar each other to the subjects. An analysis of the responses given by subjects in this condition showed that participants were strongly biased towards the two middle categories: the 15 mm ball and the 20 mm ball.

Conclusions

- Results of the experiment show that it is possible to evaluate the size of the less vibrating object in the physical interaction.

- *High order structure* and frequency content seems marginally important for the subjective evaluation of the size of the less sounding object.

- The size of the less sounding object seems to be carried by the amplitude domain.

4.5 Scaling with real sounds

The previous experiment demonstrated that we can categorise the size of the exciter. Goal of this second research was to investigate whether listeners were able to provide metrical estimations of the size of the exciter as well as in the researches investigating the size of the sounding object [49, 142]. Furthermore, goal of the research was to investigate whether listeners could evaluate the size of the exciter with no foregoing information about the physical event. Moreover, aim of the research was to test once again the framework presented in the opening section. The acoustic parameters highlighted by the framework were evaluated as possible predictors of listeners' performance.

The study was divided in two experiments. In the first experiment listeners evaluated metrically the size of the ball from the sound it produced when dropped onto a plate. In the second experiment listeners performed the same scaling but balls could be dropped upon two plates different in diameter.

4.5.1 Experiment 1

Method

Apparatus Seven solid wooden balls (pine) of 10, 15, 20, 25, 30, 40, 50 mm of diameter and weighing, respectively, 0.35, 1, 2.9, 4.6, 8.6, 22.2, 44.5 g, and a baked clay plate of 215 mm of diameter were used for the experiment. Balls were dropped in the middle of the plate and from a height of 150 mm. Listeners sat facing the apparatus. The entire apparatus and the experimenter were two meters away from the listener and they were occluded from the listener's view by a 1.5×1.5 m frame covered with opaque paper.

Procedure Ten undergraduates of the University of Padova volunteered for the experiment. They all reported having a normal hearing. Before the session started, the experimenter dropped a ball, randomly selected from the ball set. Then, the experimenter asked the listener if he/she could tell the shape of the object that had just been dropped. The experiment then began. Within the experimental session each trial (single ball-size dropped on the plate) was repeated five times in a nested randomised order. Therefore, thirty-five trials were presented to each subject. On each trial the experimenter dropped the same ball three times following subject request. During this time listeners had to draw and modify a circle on a computer screen as large as the ball they thought had just been dropped by mean of a custom software. At the end of the experiment, as in the successive one, participants were asked questions about the shape and material of the impacting object (the ball) and what surface, shape and material, it was dropped on. These questions were intended to determine what overall knowledge the participants had obtained. Furthermore, in this experiment, as in the successive one, listeners received no foregoing information about either the size of the balls, the height from which the balls were dropped or the material of the balls and the plate.

Results

As far as the preliminary question is concerned, all participants answered that a spherical object had been dropped. Average circle diameters, in millimetres, were calculated for each ball size and each subject. These averages are reported in Figure 4.2. An analysis of variance performed on the data showed that listeners discriminated between the different sizes of the balls.

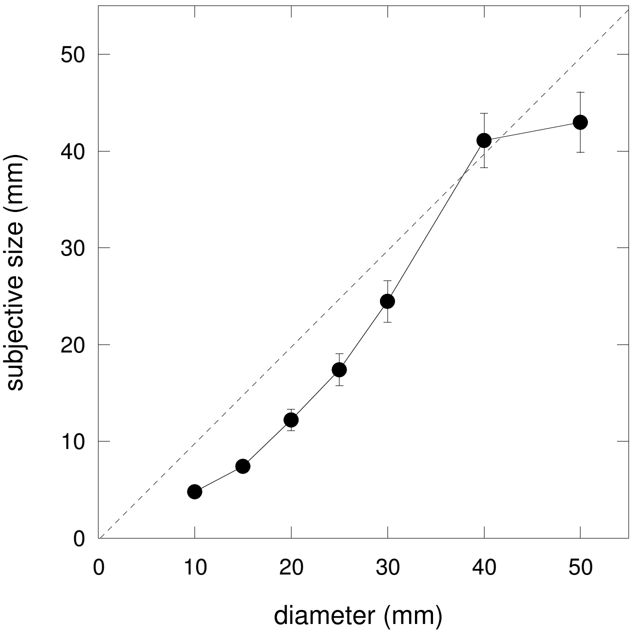


Figure 4.2: Subjective size as a function of the actual size of the ball. Vertical bars represent the standard error of the mean. The diagonal line represents a linear function with origin in zero and slope of unity.

4.5.2 Experiment 2

Method

Apparatus and procedure Ten undergraduates of the University of Padova volunteered for the experiment. They all reported having a normal hearing. None of the subjects had participated in the previous experiment. The procedure was the same as in experiment one. A second baked clay plate of 185 mm of diameter was used in addition to the 215 mm diameter plate from experiment one. This second plate had the same characteristics of the plate used in the previous experiment (material, shape etc.) and was only different in diameter. As in the first experiment, before the experimental session, listeners heard the sound of a single ball, then they were asked if they could recognise the shape of the impacting object from its sound. The procedure was the same as in the first experiment. However, the sequence of trials was extended with the thirty-five new trials obtained dropping the balls also on the 185 mm plate. This resulted in a total of seventy trials per experimental session.

Results and discussion

As far as the preliminary question is concerned, all participants answered that a spherical object had been dropped. The average circle size (in mm) was calculated for each stimulus and each subject. An analysis of variance was performed on the data. Listeners discriminated well the

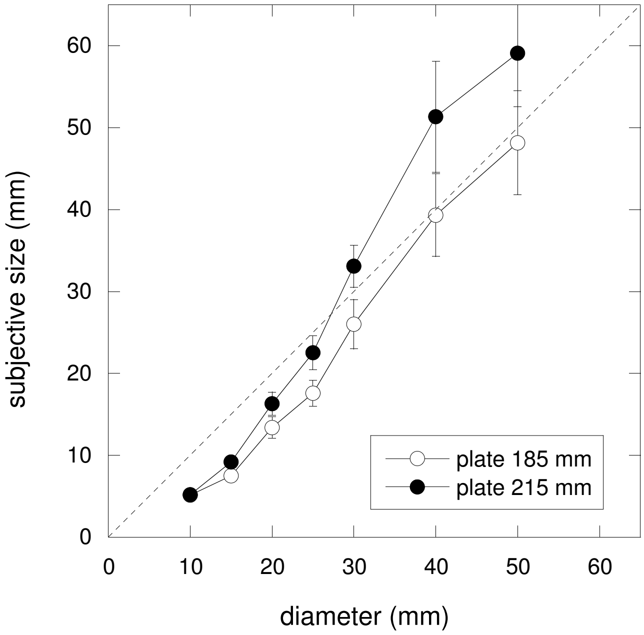


Figure 4.3: Subjective size as a function of the actual size of the ball. White dots line shows the subjective size of the balls when balls were dropped upon the smallest plate. Black dots shows the subjective size of the balls when balls were dropped upon the largest plate. Vertical bars represent the standard error of the mean. The diagonal line represents a linear function with origin in zero and slope of unity.

sizes of the balls, however, the size of the balls dropped onto the small plate was underestimated compared to the size of the same balls when dropped onto the large plate (see Figure 4.3). In particular, largest balls were perceived larger when dropped onto the 215 mm diameter plate. On the contrary, smallest balls had similar subjective size when dropped onto the small plate or the large plate. For example, the 10 mm ball was perceived 5 mm large either when dropped onto the 185 mm diameter plate or onto the 215 mm diameter plate. On the contrary, the 50 mm ball was perceived as 48 mm large when dropped onto the 185 mm diameter plate but more than one centimetre larger when dropped onto the 215 mm diameter plate (see Figure 4.3).

Participants reported that balls and plates could be made of more than one material: metal, ceramic, glass or wooden balls (one listener); metal or ceramic plates. However, none of the subjects reported that plates had different diameters.

4.5.3 Relationship between subjective estimation and acoustic signal

At the end of the experiment the sounds produced by dropping the balls upon the 215 mm diameter plate and upon the 185 mm diameter plate were recorded and stored into a computer hard drive. The following acoustic parameters were computed for each sound file: the duration, the amplitude peak, the spectral centroid, the average RMS power and the time between the first and the second bounce. These values, together with the inertial mass of the balls and their diameters, were then used to compute a stepwise multiple linear regression in order to understand which parameter could predict the performance obtained in experiments.

Analysis of experiment one

The best predictor for the performance of experiment one was the actual diameter of the ball. However, diameter per se can not affect neither the physical event nor the acoustical event. For this reason diameter was excluded from the analysis. Within the remaining predictors, the power of the signal was the best. Furthermore, a more complex model including power, duration and peak amplitude of the signal showed the best fit.

Analysis of experiment two

The actual diameter of the balls dropped onto 215 mm diameter plate predicted well the perceptual size of the balls. Excluding this predictor, within the remaining predictors, the power of the signal was the best. Furthermore, a model including power, duration and centroid of the signal showed the best fit.

The actual diameter of the balls dropped onto the 185 mm diameter plate predicted well the perceptual size of the balls. By excluding this predictor, the best of remaining predictor was the power of the signal. Furthermore, a model including together the power of the signal and mass of the ball showed the best fit.

A second stepwise multiple linear regression was performed. This analysis was performed in order to understand which predictor could explain the difference in the subjective sizes of the largest balls when dropped onto the two plates (see Figure 4.3). The difference between the subjective sizes of the balls when dropped onto the 215 mm plate with the subjective sizes of the balls when dropped onto the 185 mm plate was calculated for each ball size. The same differences were calculated for the correspondent acoustical indexes. All differences were used, respectively, as a dependent variable and predictors for the new multiple stepwise regression. The diameter and the mass of the ball were excluded from this analysis since their values do not change by changing the plate. The increasing difference between the centroids given by the sounds of the two plates when increasing the size of the ball could explain the difference in the subjective size.

4.5.4 Discussion

In the first experiment listeners scaled correctly the size of the balls. Furthermore, the size of the balls dropped was slightly underestimated compared to their actual size (see Figure 4.2). Also in the second experiment listeners were able to perform a good scaling of the size of the balls. However, the subjective size of the ball was dependent on the plate which the ball was dropped upon. This effect was most evident for the largest balls. Compared to their actual size, the perceptual size of the balls was slightly underestimated when balls were dropped on the smallest plate and both underestimated and overestimated when balls were dropped on the largest plate (see Figure 4.3). On a closer inspection results show a larger variability than the data of the first experiment (compare standard errors in Figure 4.2 and 4.3), especially for the largest balls. In fact, while in the first experiment results showed a strong consistency across listeners, with all of them estimating similar perceptual sizes, in the second experiment listeners estimations were spread over a larger range of values.

Results of both experiments demonstrated that subjects were able to estimate the size of the exciter from the sound it produced impacting upon the resonator. Furthermore, experiment two showed that the subjective size was biased by resonator: largest balls were judged slightly larger when dropped on the 215 mm diameter plate. Moreover, the power of the sound was the most efficient predictor in order to scale the size of the balls. This result corroborates the findings of the previous experiment where the manipulation of this parameter was the most effective in impairing the performance in the categorisation task. This finding validates, once again, the suitability of the framework presented in the opening section.

The metrical estimation of the size of the ball was dependent on the plate which the ball was dropped upon. An analysis of centroid values shows that the higher the mass impacting the plates the higher the difference in the centroids. In addition, independently by the mass of the ball, the sound of both plates was characterised by an evident pitch, with the pitch of the smallest plate being the highest. As a consequence, independently from the plate, all sounds produced by the lightest balls were characterised by a bright quality. On the contrary, the heavier the mass impacting the plates the more evident the difference in timbre of the sounds produced by the two plates. In conclusion, listeners were addressing a characteristic due to the plate dimension, the pitch and its effect on the overall brightness of the sound, to the size of the ball.

Participants reports after the experiment show that they were able to recognise the sound source event: all listeners reported they heard balls dropped onto plates. As pointed out in the previous sections, impact sounds produced by balls are clearly distinguishable from other events because of their *high order structure* [251]. Furthermore, Kunkler-Peck and Turvey [142] already demonstrated that we can recognise the shape of a sounding object (the plate) from its sound. A certain confusion arose when listeners had to guess the material of the two objects, especially the exciter. Across the experiments only one subject reported balls were made of wood, all the rest thought that balls were made of metal, glass or ceramic.

4.5.5 Conclusion

- Results of the experiments show that it is possible to evaluate metrically the size of the less vibrating object even with no prior knowledge of the physical interaction.

- The size of the less sounding object can be affected by the sounding object.

- The power of the sound is the most powerful predictor for describing listeners' performance.

4.6 Interaction between height, distance and mass

Previous researches demonstrated that the manipulation of the power of the sound was the most effective in impairing listeners' performance in the categorisation task (research 4.4). Furthermore, the power of the sound was the best predictor in order to describe listeners' performance with the metrical scaling task (research 4.5). Therefore, according to our results, power conveys information about the physical properties of the exciter. However, in a given impact, we can manipulate the power of the resulting sound in three ways: by manipulating the mass m of the object that is impacting the sounding object; by manipulating the height h from which the object is dropped; by manipulating the distance (d) between the sound source event and the listeners.

In some of the studies in ecological acoustics authors claimed that listeners perceive directly the sound source event (the distal stimulus) [92, 95, 97, 79, 80, 240] rather than perceive the proximal stimulus. Consequently, if we hypothesise direct perception of the distal stimulus, we should expect that any manipulation in the distal stimulus should be perceived separately and independently from any other manipulation.

The aim of the current research was to investigate whether direct perception of the physical characteristics of the sound source event was possible. In fact, in the experiment, listeners will evaluate sounds produced by different distal stimuli whose resulting sounds (the proximal stimuli) should be, in principle, similar. In particular, aim of the current research was investigated whether the physical variations of mass (m), height (h), and listener's distance from the sound source (d) influenced the estimations of subjective weight (M), subjective height (H), and subjective distance of the sound source (D).

Method

Stimuli and apparatus Sounds produced by three solid steel balls weighting, respectively, 6, 12 and 24 g, and dropped upon a wooden board of 1500 × 500 × 20 (height) mm from, respectively, 100, 200 and 400 mm were recorded in a quiet room. Recordings were made positioning the microphone at three distances: 200, 400 and 800 mm. This resulted in a total of twenty-seven recordings. Recordings were made using a MKH20 Sennheiser microphone and a portable Tascam DA-P1 DAT recorder. Successively recordings were stored in the hard drive of a Pentium computer. The sound files were coded at 44.1 kHz sample rate and 16 bits resolution. During the experiment, sound files were presented to the listener by using a custom software that also controlled the experiment. The monophonic audio output of a Yamaha sound card was presented through Sennheiser HD414 headphones.

Procedure Ten listeners, who reported having normal hearing, participated individually in the experiment. The magnitude estimation method was used. With this method, subjects have to estimate numerically their sensations. In particular, they have assigned a number that corresponds to the magnitude of their sensation. During each trial subjects listened to a sound, then they had to manipulate on the computer screen three horizontal scroll bars, for evaluating, respectively, the mass of the ball (m), the height from which the ball was dropped (h), and the distance of the sound source event (d). Scroll bars ranged between zero (minimum subjective mass, height or distance) and one-hundred (maximum subjective mass, height or distance). At the beginning of each trial all scroll bars were positioned on the left origin (zero). The order of stimuli was randomized for each subject. Each recording was presented four times during the experiment. Therefore, a total of one-hundred and eight trials were run by each subject. Before the experimental session, subjects were informed that they were going the listen to sounds produced by steel balls dropped upon a wooden board. No further information was given to the subjects.

Results

Listeners estimations did not cover all the possible 100 values permitted by the scroll bars. On the contrary, estimations covered a range between ∼10 and ∼90. For this reason, listeners' estimations were transformed according to the following equation:

$$s_r = 100 \times \left(\frac{s - s_{min}}{s_{max} - s_{min}} \right) \quad , \tag{4.9}$$

where s_r is the resulting score, s is the mean score given by the subject for a specific stimulus, s_{min} is the minimum score across subjects within a specific subjective estimation (mass, height or distance) set and s_{max} is the maximum score across subjects within a specific subjective estimation (mass, height or distance) set. This transformation do not affect the rank order of the estimations and expands the responses' scale up to the minimum/maximum permitted value.

Estimation of subjective mass A 3 (masses) × 3 (heights) × 3 (distances) analysis of variance was performed on the average subjects' mass estimations for each stimulus (see Figure 4.4). Listeners scaled correctly the mass of the balls: the higher the mass of the ball the higher listeners' estimation: $F(2, 18) = 2675.9,\ p < .0001$. Furthermore, the estimation of the subjective mass was dependent on the distance of the sound source event: $F(2, 18) = 1300.49,\ p < .0001$. The higher the physical distance, the smaller the subjective mass. Also the height showed an effect on the perceived mass: $F(2, 18) = 36.09,\ p < .0001$. Listeners estimated balls lighter when they were dropped from elevated heights.

Estimation of subjective distance A 3 (distances) × 3 (heights) ×3 (masses) analysis of variance was performed on the average subjects' estimations of distance for each stimulus (see Figure 4.5). Listeners scaled correctly the distances of the recordings: the higher the distance, the higher listeners' estimation: $F(2, 18) = 2140.84,\ p < .0001$. Furthermore, the estimation of the subjective distance of the sound source event was dependent on the mass: $F(2, 18) = 224.77,\ p < .0001$. The higher the mass of the ball, the smaller the subjective distance of the sound event. Also the height showed an effect on the perceived distance: $F(2, 18) = 5.69,\ p = .012$. Listeners estimated balls nearer when they were dropped from higher heights.

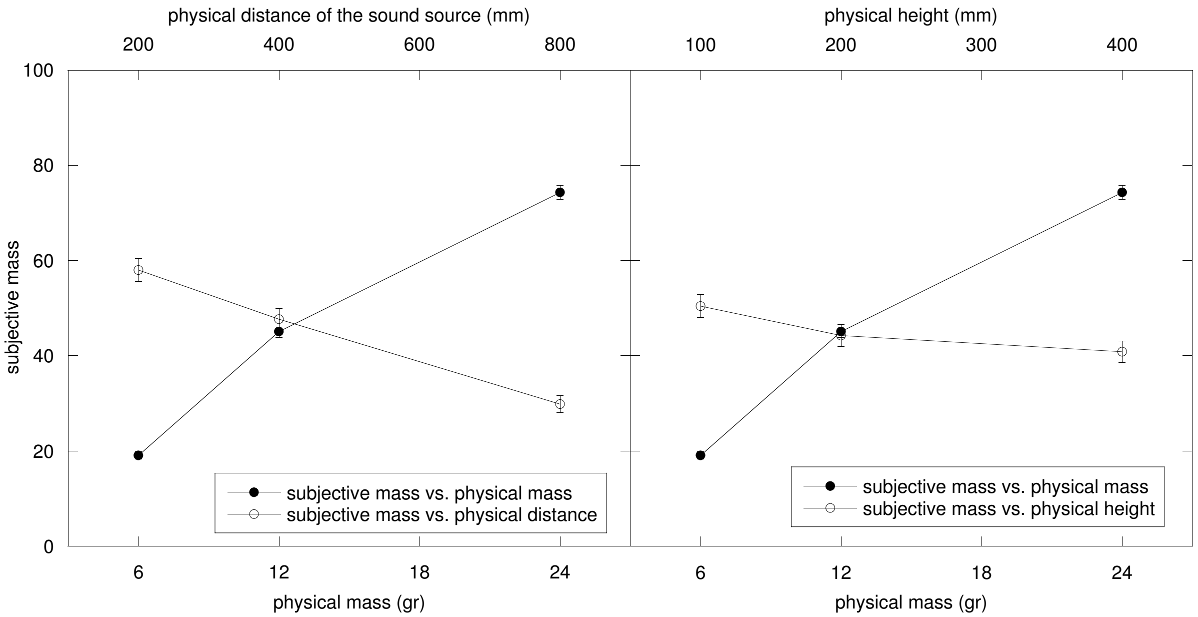


Figure 4.4: On the left panel: subjective mass as a function of the masses of the balls and as a function of the distances of the sound source event. On the right panel: subjective mass as a function of the masses of the balls and as a function of the height from which balls were dropped. Vertical bars represent the standard error of the mean.

Estimation of subjective height A 3 (heights) × 3 (masses) × 3 (distances) analysis of variance was performed on the average subjects' estimations of height for each stimulus (see Figure 4.6). Listeners scaled correctly the height: the higher the height from which the ball was dropped, the higher listeners' estimation: $F(2, 18) = 3415.80,\ p < .0001$. Furthermore, the subjective height was dependent on the mass: $F(2, 18) = 700.45,\ p < .0001$. The greater the mass of the ball, the smaller the subjective height. Also the physical distance of the sound source event showed an effect on the perceived falling height: $F(2, 18) = 8.55,\ p = .002$. Listeners estimated the height smaller when balls were dropped from longer distance.

Discussion

The results obtained show that any manipulation of the sound source event affects each subjective estimation (M, D and H). Therefore, listeners are not able to extract from the sound the information concerning the sole parameter they are asked to evaluate. In particular, heavy balls are not only balls with high masses but also balls that are dropped close to the listener and balls that are dropped at elevated heights. At the same time, distant sound source events are not only those events far from the listeners but also those obtained by dropping light balls from a moderate height. Furthermore, balls heard as dropped from elevated heights are not only those dropped from elevated heights but also heavy balls, dropped close to the listener.

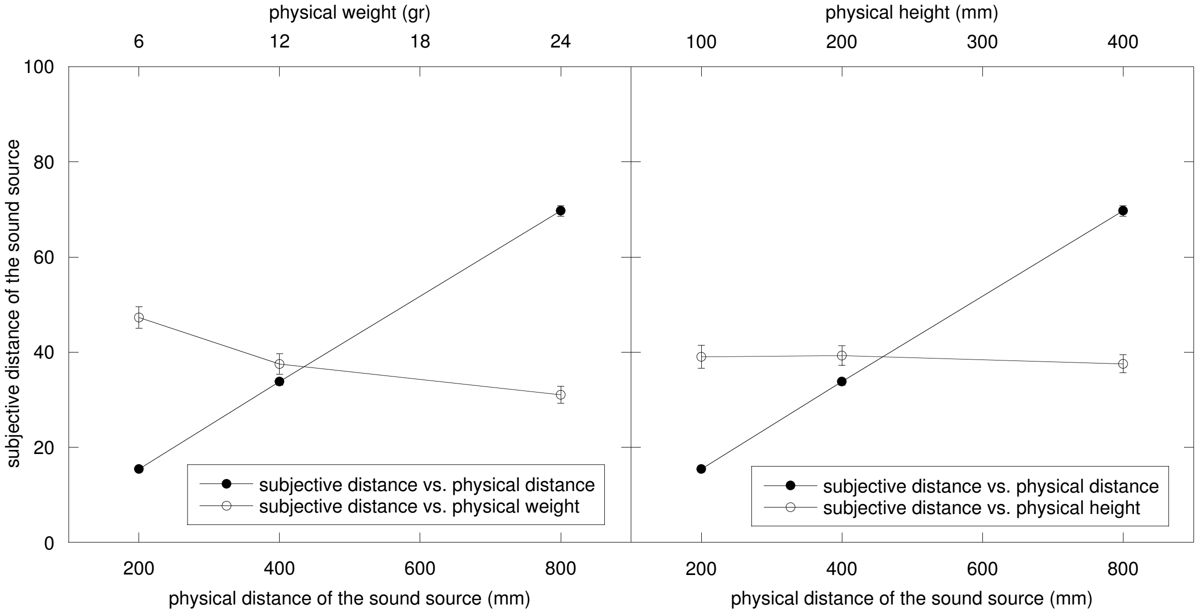


Figure 4.5: On the left panel: subjective distance as a function of the distances of the sound source event and as a function of the masses of the balls. On the right panel: subjective distance as a function of the distances of the sound source event and as a function of the heights from which balls where dropped. Vertical bars represent the standard error of the mean.

4.6.1 Conclusions

- A large increment in the mass (m) of the impacting ball corresponds to a large increment in the perceived mass (M), a large decrement in the perceived distance and a large decrement in the perceived height (H).

- A large increment in the physical distance (d) from the sound source event corresponds to a large increment in the perceived distance (D), a moderate decrement in the perceived mass (M) and a small decrement in the perceived height (H).

- A large increment in the physical height (h) corresponds to a large decrement in the perceived height (H), a moderate decrement in the perceived mass (M) and a small decrement in the perceived distance (D).

4.7 General discussion

Results of the current researches provide new evidences for the perception of physical properties of objects from sound. In particular, results demonstrated that physical characteristics of the less sounding object (i.e. size, weight, distance and height) can be perceived.

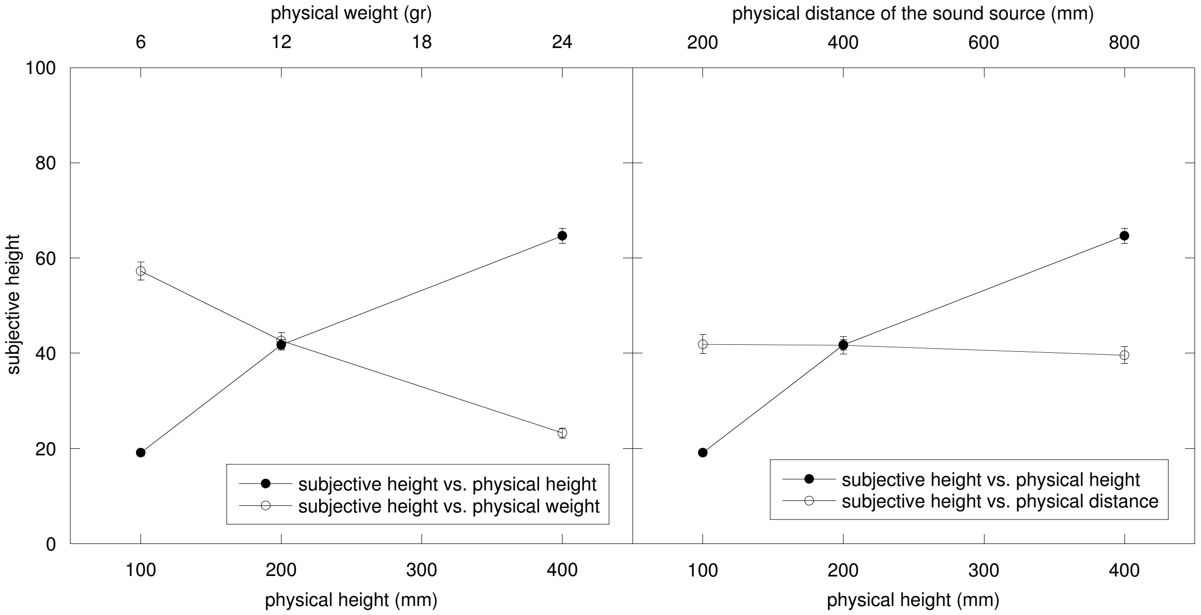


Figure 4.6: On the left panel: subjective height as a function of the heights from which balls where dropped and as a function of the masses of the balls. On the right panel: subjective height as a function of the heights from which balls where dropped and as a function of the distances of the sound source event. Vertical bars represent the standard error of the mean.

Findings of the first research demonstrated that listeners can evaluate the size of a ball from the sound it produces impacting upon a plate in a categorisation task. Furthermore, this ability was independent from the plate which the ball was dropped on. All manipulations performed on the original sounds produced an impairment in the performance. In particular, the manipulation of the *high order structure* decreased listeners' performance by a constant proportion. Also the manipulation of the frequency content of the sounds produced a linear decrement in listeners' performance, although, the impairment was larger than with the previous manipulation. The manipulation of the power of the sound, instead, was the most effective in impairing listeners' performance and, moreover, to only affecting listeners' response pattern. This result confirms the prediction provided by the framework proposed in the opening section of this chapter: in a physical interaction the exciter is mainly providing the energy for the vibration of the sounding object and, consequently, its effect on the resulting sound concerns the amplitude domain.

Findings of the second research demonstrated the listeners can provide metrical and veridical estimations of the size of the exciter. However, such estimations can be biased by the interaction between the exciter and the sounding object. In fact, in the second experiment of this research, balls were judged larger when dropped onto the largest plate. This result contradicts the findings of the first research. However, in the two researches, tasks required from subjects were substantially different (categorisation vs. metrical estimation). Furthermore, in the first research listeners knew that balls were dropped upon three different plates. On the contrary, in the second research, listeners had no foregoing information about the physical event. In addition, in the interviews after the experiment, they reported that, likely, only one plate was used in the experiment. Regressions between acoustical, physical predictors and subjective size provided further corroboration for the framework proposed in the opening section. However, the best predictor for listeners' estimations was the actual size of the ball. Therefore, in this research, listeners performances ware slightly better than any single acoustical or physical parameter could predict. As a consequence, it is possible that subjects use more than one cue, either acoustical or physical, for evaluating the size of the object and providing the estimations.

Findings of the third research demonstrated that listeners cannot evaluate independently a single physical variation in a sound source event when multiple manipulations of the physical interaction are performed. In fact, if estimations were strictly dependent on the single physical variation listeners were asked to evaluate, estimations for this physical parameter had to be independent from the manipulation of the others. For example, the estimation of the mass had to be unaffected by the distance from which the event happened and the height from which the ball was dropped. On the contrary, any manipulation of the physical situation corresponded to more than one variation in the subjective evaluations. In the experiment, when listeners had to evaluate the mass of the ball, estimations were dependent not only on the mass of the ball but also on the distance of the event and on the height from which the ball was dropped. *Mutatis mutandis* such dependencies have been found for any of the remaining subjective estimations: distance and height.

4.7.1 Distal stimulus vs. proximal stimulus

One of the debates in ecological acoustics is about direct perception. According to some author when we listen to a sound we do not listen to its acoustical characteristics (i.e. frequency, amplitude content, etc.) [92, 95, 97, 79, 80, 240]. On the contrary, we perceive directly the sound source event. In the most orthodox and caricatural version direct perception would correspond to a 'photograph' of the sound producing event. However, researches found partial evidences for a one to one correspondence between the sound source event and listeners' estimations.

According to Gaver [95, 97] we can distinguish between *everyday listening* and *musical listening*. The first corresponds to the direct perception of the sound source event (the distal stimulus), the second corresponds to the perception of the proximal stimulus. *Musical listening* would happen when *everyday listening* cannot occur, therefore, in those situations where it is not possible to recognise, either correctly or incorrectly, the sound source event [95, 97]. Nonetheless, results of the current researches demonstrate that subjects: recognise the sound source event (*musical listening*); they are biased by the distal stimulus (for example in the second and third research).

Overall, direct perception seems to apply for the recognition of the sound source event per se (for example rolling vs. bouncing) and not (or not as much) for the characteristics of objects that produce the sound source event (for example, mass, size, etc.). Moreover, in order to investigate whether the direct perception applies to static properties of objects, experiments need to manipulate more than one physical characteristic at time (as in the second and third research). In this way we can investigate dependencies and independencies between physical features of the sound source event and subjective evaluations and, consequently, if we perceive directly the properties of the objects involved in the sound source event or not.

4.7.2 Sounding object vs. non sounding object

In studies on impact sounds both the perception of the properties of the sounding object [49, 142, 145] and the exciter [82] have been investigated. In two of the researches investigating the sounding object [145, 142] authors indicated the frequency domain as responsible for carrying information about the ratio between the height and the width of a stuck bar [145] and the information about the shape of a plate [142]. In fact, the vibrational pattern of any object is dependent on its physical dimension and shape. As a consequence, any manipulation of the metrical dimensions of the sounding object corresponds to a different frequency content in the resulting sound.

In the experiments reported in this chapter we studied the perception of the non sounding object of the physical impact: the object that provides to the sounding object the necessary energy for the vibration. In an ideal interaction where the non sounding object does not vibrate, a variation of its impacting force corresponds mainly to a variation in the amplitude content of the sound produced by the sounding object. Of course, in interactions between sounding and non sounding objects, both objects vibrate but the second vibrates less than the first and, consequently, the statement above remains valid: a manipulation concerning the non sounding object corresponds to a variation in the amplitude domain. Consequently, good predictors for listeners' estimations can be found in the amplitude domain.

Overall, results obtained so far, highlight a partition between frequency and amplitude domain on the one side and sounding and non sounding object on the other side. What research has to investigate is whether listeners are aware of partition to some extent: whether listeners posses knowledge of the sound source event and its effect on the resulting sound.

4.7.3 A problem of segregation

Results of the current researches and previous researches rise a problem: how do listeners separate the information concerning the sounding object from the information concerning the less sounding object? In fact, the sound reaching our ears is one. However, this sound is the result of the interaction between (at least) two objects: the sounding object and the exciter. Furthermore, listeners demonstrated that they are able to scale the physical properties of the former [142, 49] and the latter [82] (see also the second and third research in the current chapter).

It is possible that listeners posses a mental representation of the physical event and such representation is shaped according to a physical model of the sound producing event. As a consequence, listeners would act: by selecting the representation that is appropriate for the physical interaction of the sound producing event (this selection can be performed extracting the *high order structure* of the sound); by evaluating the physical model and, in particular, the role of the object they are asked to evaluate and the consequences of its manipulation on the resulting sound; by selecting in the resulting sound the acoustic features that are most likely to be related to the object they are asked to evaluate. Although the existence of a mental representation of the physical interaction and its predictions on the resulting sound seems an intriguing hypothesis, so far, literature provided no explicit evidence about it.

Chapter 5

Material categorization and hardness scaling in real and synthetic impact sounds

Bruno L. Giordano
Università di Udine – Faculty of Education
Udine, Italy
Università di Padova – Department of General Psychology
Padova, Italy
bruno.giordano@unipd.it

5.1 Introduction

Several studies demonstrated the ability to scale or discriminate properly different features of sound sources on the basis of auditory information alone. Stunning abilities have been found in the judgments of geometrical properties of the sound source [145, 49, 142] or in judgments of geometry–independent features [82]. It was found that listeners are able to categorize correctly even more complex features, such as the type of interaction between objects [251], or the gender of walkers from the sound of their footsteps [153].

The ecological approach explains these abilities by assuming a direct link between the physical features of the sound source (distal stimulus) and the perceptual level. Direct and non mediated perception of the sound source features would be possible in virtue of the fact that the acoustical signal (proximal stimulus) specifies richly and uniquely the sound source [50]. The detection of a specific property of the physical event is hypothesized to be based on so-called invariants, structural properties of the acoustical signal that specify a given property of an event despite variations in other features of the sound source [50]. Invariant properties of the proximal stimulus are, for example, those that allow to recognize a piano despite the drastic acoustical signal variations associated, for example, with changes in the environment, in pitch and dynamics.

Theories based on the notion of invariants imply a one-to-one mapping between the physical features of the sound source and higher level acoustical properties (i.e., *specificity*). Furthermore, they imply the ability of listeners to scale or categorize a given feature *veridically*, and *independently* from variations in extraneous features of the sound source. The empirical test of these assumptions requires the use of *perturbation variables* 1. If perturbation variables have significant effects on recognition, then the assumptions of veridicality, and independence must be discarded for the investigated sound source feature.

These assumptions were directly tested in a first group of experiments, which investigated material type categorization with real impact sounds. An indirect test to these assumptions was provided by another set of experiments, that investigated hardness scaling with synthetic sounds. All stimuli were generated by the impact of a highly damped object, whose vibrations decay rapidly after impact (*hammer* or *non-sounding object*), on a resonating object, whose vibrations are the main source of fluctuation in the pressure of the air medium (*sounding object*). In the first set of experiments, hammer properties were treated as perturbation variables. In the second set of experiments, hammer properties were treated both as perturbation variables and as object of direct investigation. In both studies, geometrical properties of the sounding objects were used as perturbation variables.

Finally, both series of experiments addressed a more methodological problem, related to the nature of the investigated stimuli. Two options are actually available: real sounds, generated by manipulating real physical objects, or synthetic sounds, generated by manipulating simulations of the mechanical behavior of the physical objects. By definition, a model is a simplification of the modelled object. As such it carries, potentially, some simplifications. In particular, if physical models for sound source recognition research are concerned, there is the possibility that sound models cut out relevant acoustical information for the detection of a particular feature of the source. This suspicion was raised by Carello et al. [50] in reviewing Lutfi et al. [163] results on material discrimination in synthetic sounds. Performance was found far from perfect. On the contrary, a different study [142], conducted with real sounds, demonstrated almost perfect levels of recognition. Thus the legitimate objection was that the difference in results could be attributed to the fact that part of the acoustical information relevant to material categorization or discrimination, was present in the real stimuli investigated by Kunkler-Peck at al. [142], but not in those investigated by Lutfi et al. [163]. Results of previous researches will be compared with results reviewed in this chapter. This will allow us to finally develop a simple criterion for assessing the perceptual validity of a synthesis model.

5.2 Material

5.2.1 Previous researches

Several researches examined the potential acoustic information available for material recovery from impact sounds. Wildes and Richards [253] developed an analysis of the mechanical behavior of solids, in order to find an acoustical measure that uniquely identified material, despite variations in geometrical features. From the physical point of view materials can be characterized using the coefficient of internal friction $\tan\phi$, which measures their degree of anelasticity. This physical measure is measurable using the decay time of vibration or, alternatively, its bandwidth. This relation is expressed by Eq. (5.1).

$$\tan\phi = \frac{1}{\pi f t_e} = Q^{-1} \quad , \tag{5.1}$$

where f is the frequency of the signal, t_e is the time required for amplitude to decrease to $1/e$ of its initial value, and Q^{-1} is the bandwidth of the spectral components. In ascending order of $\tan\phi$, we have rubber, wood, glass, and metals. From rubber to metals spectral components have progressively longer decay times, and progressively decreasing bandwidths. As the invariant for material type is based on the acoustical measurement of the coefficient of internal friction, one should expect this physical measure to remain constant across variations of the shape of the objects. On the contrary, Wert [252] showed that the shape independence of the $\tan\phi$ coefficient is only an approximation. Furthermore Wildes and Richards model, assumes a simple relation of inverse proportionality between frequency and decay time. This was found to be a simplification, as measurements conducted on struck bars and plates sounds found the relationship between the frequency and decay times of the spectral components to be quadratic [92] or more complex than quadratic [52, 53].

If we focus on simple acoustical properties, the ambiguity of the acoustical signal features in respect to mechanical material properties emerges. Lutfi et al. [163], in solving the equation for the motion of the struck-clamped bar, outlined that both geometrical and non geometrical features of the struck bars influence the amplitude, frequency and decay times of the spectral components.

The ambiguity of the acoustical features in respect to material type is in contrast with previous researches on material recognition in real impact sounds. Gaver [92] tested material recognition in sounds generated by percussing wood and iron bars of different lengths, with percent correct performances between 96% and 99%. Kunkler-Peck et al. [142] investigated shape and material recognition in struck plates sounds. Material recognition was almost perfect.

Results gathered on synthetic stimuli provide insight on the acoustical determinants of material recognition and discrimination. Lutfi et al. [163] studied material discrimination in synthetic struck clamped bar sounds. Synthesis parameters were chosen to model glass, crystal, quartz, and different metals. Performance was analyzed in respect to three acoustical features: amplitude, decay rate and frequency. This analysis revealed that material discrimination was mainly based on frequency. Amplitude and decay rate had only a secondary role. Klatzky et al. [137] investigated material discrimination in stimuli with variable frequency and decay modulus τ_d[1]. In a first experiment subjects were asked to estimate the perceived difference in material. Results indicated that judgments were significantly influenced by both τ_d and frequency, even though the contribution of the first to judgments was higher than the latter. The same experiment, conducted on amplitude equalized signals, did not give different results, thus pointing toward an absence of this acoustical

[1] $\tau_d = 1/\pi \tan\phi$ and $t_e = \tau_d/f$.

variable on material recognition. An effect of both the τ_d coefficient and the fundamental frequency was found in a subsequent experiment, where subjects were asked to categorize material into four categories: steel, rubber, glass and plexiglass. Steel and glass were chosen for higher τ_d values than rubber and wood, and, thus, for longer decay times. Glass and wood were chosen for higher frequencies than steel and rubber. Similar results were obtained by [10]. Stimuli varied in the quality factor Q of the spectral components[2] and in frequency. Participants had to categorize material type using four response categories: steel, glass, wood, and rubber. A dependence of material categorization on both controlled acoustical parameters was found. As t_e, and thus decay times, increased, material categorization changed from rubber to wood to glass, and finally, to steel. Coherently with what found in [137], steel was chosen for lower frequencies than glass while, contrary to their result, a slight tendency to choose rubber for higher frequencies than wood was observed.

As frequency content is also determined by the geometrical features of the objects, the effects of frequency on material recognition point toward a limited ability of subjects to recognize correctly material from impact sounds. The following experiments tested these effects with real impact sounds.

5.2.2 Experiments

Three experiments tested material recognition upon the effect of different perturbation variables. Stimuli were generated by striking rectangular plates made of four different materials: steel, glass, wood, and plastic. In all the experiments the area of the plates ranged from 75 to 1200 cm^2. In the first and second experiment plates were struck using a steel pendulum. Plates varied also in height/width ratio, this latter ranging from 1 (square plate) to 5. In the first experiment plates could vibrate freely after being struck by the pendulum. In the second experiment plates were artificially damped with an equal shape and area low density plastic plate. In the last experiment square plates were struck with penduli made of four different materials: steel, glass, wood, and plastic. The starting angle of the penduli was kept fixed for all the experiments.

Stimuli were characterized using two acoustical parameters. First, a measure of the frequency of the lowest spectral component F. Second, a measure of the amplitude decay velocity T_e, defined as the time required for the spectral level to decay to $1/e$ of the attack value.

In all the experiments stimuli were presented via headphones. Subjects were asked to categorize the material of the struck objects. They were not informed about the variation in perturbation variables.

5.2.3 Freely vibrating plates - material, area, and height/width ratio variation

Categorization performance can be summarized using the concept of material macro–categories, which describes pattern of confusions among physical materials. The first macro–category includes

[2] $Q = \pi f t_e$.

steel and glass materials, the second includes plastic and wood. Confusion between the two material macro–categories was absent, glass and steel being almost never chosen in wood and plastic plates sounds, and vice versa. Confusion between materials inside the same macro–category was high: glass was confused frequently with steel, and vice versa; plastic was confused frequently with wood, and vice versa. Inside the macro–categories, responses revealed a strong effect of the area, where small glass and steel plates were identified as being made of glass, while large glass and steel plates were identified as being made of steel. The same effect was found in the plastic/wood macro–category, where plastic was chosen for large area plates, and wood was chosen for small area plates. A small group of subjects revealed an opposite response profile, wood being chosen for larger area plates, plastic being chosen for small area plates. Data were analyzed using three separate logistic regression models. The first studied the probability of choosing the steel/glass macro–category over the probability of choosing the wood/plastic one. The second studied the probability of choosing steel over the probability of choosing glass. The third studied the probability of choosing wood over the probability of choosing plastic. In general identification correctness was above chance level. Recognition of the material macro–categories was almost perfect. In contrast recognition inside the macro–categories was almost at chance level. In none of the cases the height/width ratio significantly influenced the response proportions. Macro–categorical categorization was determined only by the physical material. In particular the macro–categorical response proportions did not differ between glass and steel on one side, and between wood and plastic on the other side. The steel/glass categorization was, instead, significantly influenced by both area and material. The plastic/wood categorization was determined by area variations alone.

Both F and T_e were found to have significant effects in all the regression models. The effect of the decay time T_e, on macro–categorical responses, was greater than that of F. Steel and glass were chosen for higher frequencies and longer decay times than plastic and wood. Categorization inside the macro–categories was, instead, influenced more by frequency than decay times. In particular glass was chosen for higher frequencies and shorter decay times than steel, and plastic was chosen for lower frequencies and longer decay times than wood. The absence of a significant effect of the height/width ratio variable was consistent with the absence of significant F and T_e differences associated with this variable. On the contrary area and material type influenced significantly both F and T_e. The orderings of material types in respect to F and T_e did not always correspond to that used by subjects to recognize material from sound. Steel and glass had, coherently with subjects acoustical response criteria, higher frequencies and longer decay times. Coherently glass had shorter decay times, but, contrary to the acoustical criterion used by subjects, lower fundamental frequencies. Plastic and wood material types were found not different in respect to both F and T_e.

Discussion

The almost perfect levels of performance observed in the identification of the material macro–category is consistent with results collected by Gaver [92] with iron and wood struck bars sounds. In his experiment the absence of a significant effect of the length of the bars on material categorization could be due to the fact that subjects were not given the possibility to use the glass and plastic response categories. Results are inconsistent with those collected by Kunkler-Peck et al. [142], where the plexiglass and wood categories were perfectly discriminated. A possible explanation of the inconsistency can be found in the procedural differences between his experiment and ours. Kunkler-Peck et al. generated stimuli live, behind an occlusion screen that precluded sight of the device by participants. Reverberation inside the room or additional signals generated by manipulating the struck plates to prepare trials could provide additional acoustical information than that provided to participants in our experiment.

In the current experiment macro–categorical categorization was independent from variations in the geometrical properties of the plates. Furthermore subjects responses were veridical, and were based on acoustical criteria that reflected the relationship between variations in the sound source and variations in the acoustical signals. Macro–categorical categorization was found to be based more on decay times than frequency and, consistently with results collected on synthetic stimuli [137], [10], it showed that plastic and wood were chosen for lower decay times glass and steel.

The steel versus glass and wood versus plastic categorizations, however, showed a strong effect of the geometrical properties of the sound source, as well as a drastic drop in the veridicality of responses. Plastic and wood sounds were found to be equivalent from the acoustical point of view. Likely participants relied on the only physical dimension that structured significantly the acoustical signal. This, however, does not invalidate the general effect of geometrical properties on material categorization, as it was found in the steel versus glass categorization, where these two materials differed in both F and T_e.

Subjects associated wood to small area plates, and plastic to large area plates. Coherently the first category was chosen for higher frequencies than the latter. This result is consistent with those by Klatzky et al. [137]. A smaller group of subjects showed the opposite response profile, associating plastic to higher frequencies than wood. The presence of opposite response profiles in different subjects could be one of the potential explanations for the discrepancies between Klatzky et al. and Rocchesso et al. [10] results, who observed a weak association of wood with low frequencies.

Glass was associated to smaller area plates than steel. As in the case of the plastic versus wood categorization, frequency played a greater role. The glass response was chosen for shorter decay times than steel, and consistently, glass sounds were found to have shorter decay times than steel sounds. However glass sounds had significantly lower frequencies than steel sounds, while subjects associated glass to higher frequencies than steel. Our results are consistent with those collected with synthetic sounds: both Klatzky et al. [137], as well as Rocchesso et al. [10] found glass to be associated to higher frequencies than steel. Furthermore, the fact that frequency played a greater role than decay times for categorization inside the steel ad glass macro–category aligns with results by Lutfi et al. [163].

5.2.4 Damped plates - material, area, and height/width ratio variation

External damping lead to a drop in the macro–categorical recognition performance only for glass plates, while recognition of the macro–category was almost perfect for steel, wood, and plastic sounds. Glass plates sounds were almost always categorized as belonging to the plastic and wood macro–category. The same effects of area on subjects' responses were observed with damped plates sounds. In particular the category steel was chosen for larger area steel plates than the response glass, while the response wood was chosen for smaller area glass, plastic and wood plates than the response plastic. As in the previous experiment the height/width ratio had no effect on participants responses. The steel versus glass and wood versus plastic categorizations were influenced by both the material and area of the plates. The steel versus glass categorization was significantly influenced by variations in area but not in material. The absence of an affect of the material variable is due to the fact that the glass and steel responses were almost never given for damped glass plates. The wood versus plastic categorization was, on the contrary, influenced by both area and material. As in the previous experiment, all categorization contrasts were significantly influenced by both decay times and frequency. Analysis of the macro-categorical categorization revealed, consistently with results from the previous experiment, the tendency to chose the steel or glass response for higher frequencies and longer decay times. The relative weight of F and T_e in determining subjects responses was almost equal. Analysis of the steel versus glass categorization revealed the tendency to chose the category steel for higher frequencies and longer decay times than glass. Differently from the previous experiment, in this categorization decay time played a greater role than frequency. Wood was chosen for higher frequencies and shorter decay times than plastic, where, consistently with previous experiment, frequency played a greater role in determining this categorization. Not surprisingly, external damping produced a decrease in decay times, and a non significant decrease in frequency. Damped glass plates sounds were found to have equal decay times to those of damped wood and plastic ones. Wood, plastic and glass damped plates sounds had, on the average, equal frequencies, while steel sounds had higher frequencies than all these three categories. Steel sounds had the highest decay times, glass and plastic were not different in respect to T_e, while wood damped plates sounds had the lowest decay times. Overall glass damped plates sounds were not different from the plastic ones, so that it is not surprising that they were categorized as being made of plastic or wood.

Discussion

In respect to the considered acoustical variables, damping caused glass sounds to be identical to plastic sounds. In particular, the effect of damping on T_e explains the switch in the recognized macro-category for glass sounds. The same explanation, however, does not account for the fact that steel sounds did not switch material macro-category. In fact damping caused a general decrement in the decay times, well below the values measured on freely vibrating wood and plastic plates sounds. If this parameter was the main cause for categorization within the material macro–categories, steel sounds had to be categorized as being made of wood or plastic, as happened for glass sounds. Further acoustical measurements are required to explain the different results gathered with glass and steel damped plates.

5.2.5 Freely vibrating plates - material, area, and pendulum material variation

Data revealed the same area and material effects found in the first experiment. Once again a small percentage of subjects showed an opposite response profile in respect to the plastic versus wood discrimination. Variations in the material of the percussor were found to have no effects on all the categorization contrasts. Categorization inside the material macro-categories was influenced only by the material of the plates. As in the first experiment, steel and glass were equivalent to each other in respect to the macro–categorical categorization, as well as wood and plastic. The steel versus glass categorization was influenced by area and material, while the wood versus plastic categorization was influenced by area only. Macro–categorical categorization was determined by both frequency and decay time, with the latter having the greatest weight. The glass versus steel categorization was influenced by both the considered acoustical variables, F having a greater weight than T_e. The wood versus plastic categorization, finally, was influenced only by frequency. Variations in the hammer were not associated to significant variations in F and in T_e.

Discussion

There are two alternative ways to interpret the absence of significant effects of the percussor in determining categorization of the material of the plates. The first focuses on the distal stimulus, and outlines subjects ability to distinguish variations in the features of the hammer from variations in the features of the sounding object. The second focuses on the proximal stimulus and outlines the fact that variations in the material of the pendulum were not associated to significant variations in the acoustical features. As found in previous experiments, the extent to how a perturbation variable influences the investigated sound source feature depends on its range of variation. The height/width ratio variable was not significant because it did not lead to significant variations in F and T_e, as compared to the effects associated to variations in the area. For the same reason we should expect that a larger range of variation in the material properties of the hammer would lead to changes in material categorization. This would be possible provided that the acoustical variables used by subjects to categorize material are affected by variations in the hammer material properties. The effects of the hammer on recognition of the geometry–independent properties of the sounding object was addressed with the experiments summarized in section 5.3.

5.2.6 Conclusions

Material categorization was found to be affected by the geometrical features of the sounding object. Analysis of responses revealed that veridicality depended on the level of detail in categorization. Results gathered from freely vibrating plates sounds, in fact, showed that identification of the material macro-categories was almost perfect, and was not influenced by variations in the geometrical properties of the sounding object. Accuracy of categorization inside the macro-categories dropped to chance level, and was strongly influenced by the geometrical properties of the struck objects. The dependence of categorization veridicality on the level of detail is consistent with the free identification results obtained by Gaver [92], who reported that "accuracy [of identification] depends on specificity". The orthodox ecological approach to auditory perception assumes veridicality of recognition and absence of effects of irrelevant sound source dimensions (independence). The correctness of these assumptions have been found to depend on the detail level in material categorization. Recognition of the material macro–category are in agreement with them, while categorizations inside the macro–categories disconfirm them.

An attempt to define an acoustical invariant for material type wasn't performed. However, the significant effects of the investigated perturbation variables on material categorization allow us to conclude that, although an invariant for material type may exist, it is not sufficient to completely explain material recognition. Given the limits of the coefficient of internal friction model [253], it was preferred to characterize acoustical signals using simple acoustical features. Focusing on simple properties confirmed the existence of an ambiguous relationship between objects properties and acoustical signal features [163]. Ambiguous specification of the sound source in the acoustical signal disconfirms the specificity assumption of the orthodox ecological approach. Specific acoustical criteria for material recognition were outlined. These criteria were found, when comparison was possible, consistent with all previous researches conducted using synthetic stimuli.

External damping provided a way to investigate material categorization using stimuli much more adherent to those encountered during everyday life, were externally damped objects are much more common than freely vibrating ones. Furthermore it allowed testing the effect of the decrease in decay times on material recognition via manipulation of the sound source, instead of via manipulation of the recorded signals. This sound source feature have been found to affect material categorization. In particular damped glass plates sounds were categorized within the plastic/wood macro-category. This was explained by the fact that the considered acoustical properties of damped glass plates were equal to those of the damped plastic ones.

5.3 Hardness

In this section we present two experiments concerning recognition of geometry–independent features of the sound source. In the previous experiments the focus of the investigations was on material type, investigated through categorization procedures. Here we focused on hardness, a mechanical material property, and investigated material recognition through scaling.

Scaling of the hardness of the hammer was already studied by Freed [82] with real sounds. Investigated stimuli were generated by striking four cooking pans of variable diameter with six mallets of variable hardness (metal, wood, rubber, cloth-covered wood, felt, felt-covered rubber). Analysis of the performance revealed an appropriate scaling of the hardness of the mallet, independent of the diameter of the pans.

With the present investigations we extended the research by Freed in several directions. Freed addressed hardness scaling of the hammer by using the geometrical features of the sounding object as perturbation variable. We investigated hammer hardness scaling by using both the geometrical properties of the sounding object, as well as its material as perturbation variables. Furthermore we investigated sounding object hardness scaling upon variation in the hammer properties as well as in the sounding object properties. Since all the experiments made an extensive use of perturbation variables, the specificity, independence, and veridicality assumptions of the ecological approach could be tested again. All the experiments were performed using sounds synthesized with the impact model described in chapter 8. Comparison of results presented in this section with those collected by Freed on real sounds allowed addressing again the problems connected to research on sound recognition with synthetic sounds.

5.3.1 Stimuli

Stimuli were generated by varying two parameters that modelled the properties of the sounding object, and one parameter that modelled the properties of the hammer. Sounding object properties were varied by means of the internal friction coefficient $\tan\phi$, assumed to model variations in the material of the sounding object, and by the frequency of the lowest resonant mode F, assumed to model variations in the geometrical properties of the sounding object. Hammer properties were varied through the elasticity coefficient e. A synthesis space was derived by combining a different number of equally log-spaced levels for each synthesis parameter, as described below. For each experiment a different subset of the stimuli from this synthesis space was investigated.

The $\tan\phi$ coefficient varied from 10 (t_1, corresponding to the plastic-wood macro-category in the previous experiments) to 160 (t_3, corresponding to the steel-glass macro-category in the previous experiments), the intermediate level being 40 (t_3). F was varied from 50 Hz (F_1) to 800 Hz (F_3), the intermediate level being 200 Hz (F_2). The elasticity parameter was varied from $5e^6$ (e_1) to $1e^{10}$ (e_5), the intermediate levels being $3.3e^7$ (e_2), $2.24e^8$ (e_3) and $1.495e^9$ (e_4). The lowest value of the elasticity coefficient can be conceived as representing rubber mallets, while the highest value of the elasticity coefficient can be thought as representing steel mallets.

5.3.2 Sounding object hardness

Categorization of the sounding object material was shown to be influenced by both the material and the geometrical properties of the sounding object. We expected that scaling of a material property of the sounding object to be dependent on both these modelled features of the sound source.

Methods

Ten subjects participated to the experiment. All of them reported normal hearing. Three levels for each synthesis parameter were chosen, namely F_1, F_2, F_3, t_1, t_2, t_3, e_1, e_3 and e_5. This resulted in a stimuli set that comprised 27 sounds. Subjects were told that they would have been presented several sounds generated by striking sounding objects of variable hardness with hammers of variable hardness. They were asked to rate the hardness of the sounding object on a numerical scale ranging from 1 (very soft materials) to 100 (very hard materials). Before the experiment

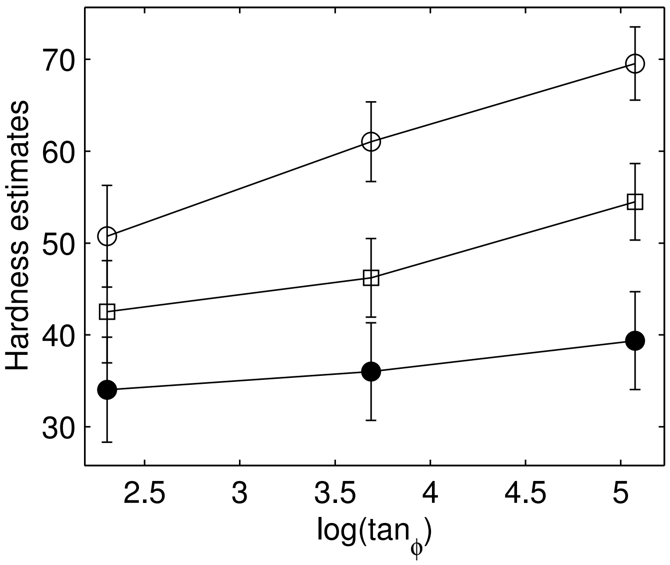


Figure 5.1: Average sounding object hardness estimates on a 1-100 scale (1 = very soft 100 = very hard), as a function of the internal friction coefficient. Filled circles: e_1; empty squares: e_3; empty circles: e_5.

started they were presented several real examples of variable hardness sounding objects struck with variable hardness mallets. Each of the 27 stimuli, presented in randomized order, was judged three times by each subject. Stimuli where presented through Sennheiser HE60 headphones connected to a Sennheiser HEV70 amplifier. The amplifier received the output signal of a Sound Blaster Live! Soundcard.

Results

Average hardness ratings were analyzed by means of a repeated measures ANOVA, with frequency, internal friction and hammer elasticity as repeated measurements factors. The 3-way interaction between the synthesis parameters, as well as the 2-way interaction between frequency and internal friction coefficient, were not significant ($F_{(8,72)} = 0.898$, $p = 0.523$; $F_{(4,36)} = 1.698$, $p = 0.172$). The interaction between the elasticity parameter and frequency, as well as that between the elasticity parameter and the coefficient of internal friction were significant ($F_{(4,36)} = 3.563$, $p = 0.015$; $F_{(4,36)} = 3.860$, $p = 0.010$). Simple effects of the coefficient of internal friction and of the elasticity parameter significantly influenced average hardness estimates ($F_{(2,18)} = 6.600$, $p = 0.007$; $F_{(2,18)} = 19.686$, $p < 0.001$). The simple effect of frequency was not significant ($F_{(2,18)} = 1.144$, $p = 0.341$).

Figure 5.1 shows the interaction between $\tan\phi$ and F, by plotting average hardness estimates as a function of $\tan\phi$, with F as factor. Figure 5.2 shows the interaction between the F and e factors, by plotting average hardness estimates as a function of the F parameter, with e as factor.

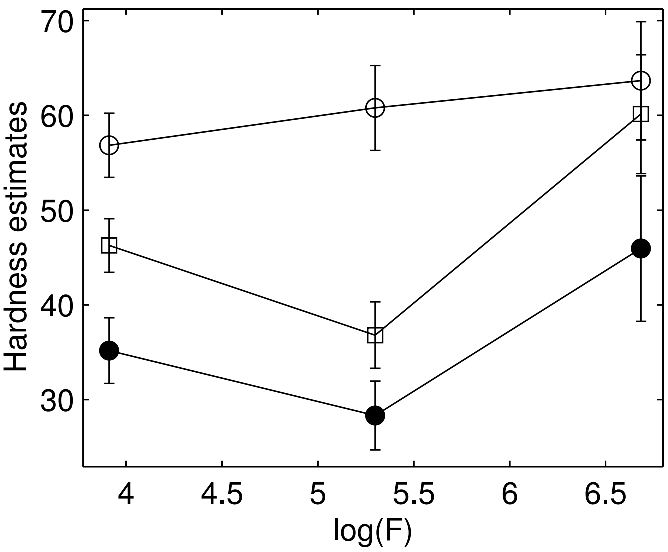


Figure 5.2: Average sounding object hardness estimates on a 1-100 scale (1 $=$ very soft 100 $=$ very hard), as a function of frequency. Filled circles: e_1; empty squares: e_3; empty circles: e_5.

Estimates of the hardness of the sounding object increased for increasing $\tan\phi$ and e. The slope of the functions that relate sounding object hardness estimates to $\tan\phi$ increases with increasing mallet hardness, so that we can conclude that the $\tan\phi$ coefficient induces higher changes in the hardness estimates for higher e values. Analysis of the interaction between e and F shows the same general effect of e on the sounding objects hardness estimates. This latter interaction has however a different origin than the one between $\tan\phi$ and F, which was due to a non parallelism of monotonic psychophysical functions. In this case it is due to a change in the shape of the psychophysical functions. In particular the effect of F on the hardness estimates is found monotonic for the highest e value used, non-monotonic for lower values of the e coefficient. In this latter case intermediate F levels (200 Hz) lead to a strong decrease of the sounding object hardness estimates.

Discussion

Materials with low $\tan\phi$ coefficient values (plastic, rubber) were estimated as the softest, while materials with high $\tan\phi$ coefficient values (steel, glass) were estimated as the hardest. As for the material categorization experiments, recovery of the material of the sounding object was found to depend on material (as modelled by $\tan\phi$) and on the geometrical properties of the sounding object (as modelled by F). These findings suggest also that the $\tan\phi$ measure alone is not sufficient to explain material recognition. The reported change of the effects of F with variable e is surprising and has no explanation. The analysis of the acoustical structure of stimuli, as well as a replication of this study with real sounds will allow outlining the causes of this effect. The significance of the effects of the elasticity coefficient points toward an inability of subjects to distinguish material properties of the mallet from material properties of the sounding object. This is particularly evident if we conceive an increase in the e coefficient as modelling hammers of increasing stiffness (i.e., from rubber to steel). When commenting the results of material type categorization with variable pendulum material, we hypothesized that a higher range in the variation of the material properties of the hammer would have led to significant changes in the recognized material. The significant effect of the e parameter on sounding object hardness estimates provides a partial confirmation to this hypothesis.

5.3.3 Hammer hardness

The final experiment investigated estimation of the hardness of the hammer upon variations in the properties of the sounding object.

Methods

Two stimuli sets were investigated. In both cases all the five levels of the e parameter were used. In the first set the $\tan\phi$ parameter was fixed to t_2, while all three levels of F were used. In the second F was kept constant at 200 Hz (F_2), while all the three levels of the $\tan\phi$ parameter were used. The combination of these synthesis parameters produced two sets of 15 stimuli. Procedure was identical to that used in the previous experiment, the only difference being that subjects were asked to rate the hardness of the hammer. All the stimuli, presented in randomized order, were judged three times by each subject. The two sets of stimuli were presented in separate sessions. The order of the sessions was counterbalanced between subjects. Ten subjects participated to the experiment. None of them participated to the previous experiment. All of them reported normal hearing.

Results

Average hardness ratings were analyzed by means of a repeated measures ANOVA. For the variable F set, average hardness ratings were significantly influenced by e ($F_{(4,36)} = 8.597$, $p < 0.001$), but not by F ($F_{(2,18)} = 2.269$, $p = 0.132$) or by its interaction with the e ($F_{(8,72)} = 1.379$, $p = 0.220$). Figure 5.3 plots the average hardness estimate as a function of e, with F as factor.

As shown, hammer hardness estimates increase with increasing e. The psychophysical functions for the different F levels appear greatly overlapped. A tendency of the hardness estimates to decrease for the highest F value (800 Hz) is observed. Contrasts were performed to test whether the estimates given for F_1 and F_2 differed significantly from those given for F_3. This additional test revealed the difference to be non significant in both cases ($F_1 - F_3$: $F_{(1,9)} = 1.903$, $p = 0.201$; $F_2 - F_3$: $F_{(1,9)} = 3.371$, $p = 0.100$;). Thus we can conclude that F did not significantly influence the hammer hardness ratings.

For the variable $\tan\phi$ set, average hardness estimates were significantly influenced by e ($F_{(4,36)} = 6.905$, $p < 0.001$) as well as by $\tan\phi$ ($F_{(2,18)} = 43.315$, $p < 0.001$). The interaction term proofed slightly significant ($F_{(8,72)} = 2.109$, $p = 0.046$). Figure 5.4 plots the average hammer hardness estimate as a function of e with $\tan\phi$ as factor.

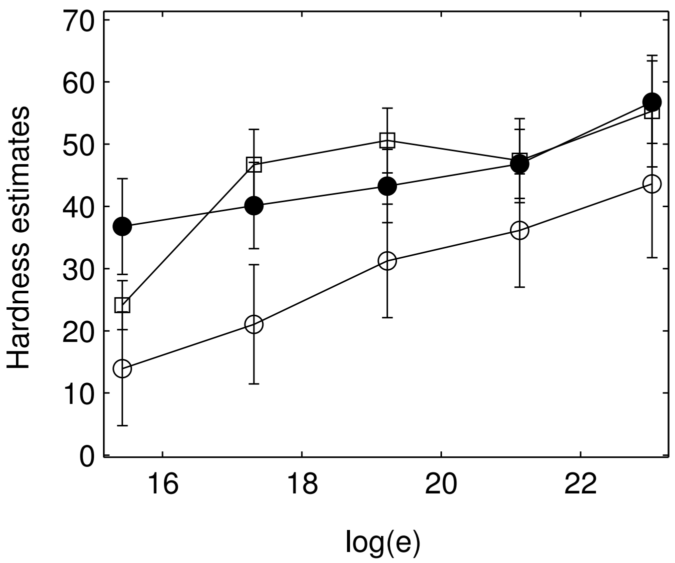


Figure 5.3: Average hammer hardness estimates on a 1-100 scale (1 = very soft 100 = very hard) for the variable F set. Filled circles: F_1; empty squares: F_2; empty circles: F_3.

As shown, hammer hardness estimates increase with increasing e and with increasing $\tan\phi$. As can bee seen in Figure 5.4 the interaction term is due to an increase in the slope of the psychophysical function for increasing $\tan\phi$ values. As observed with the previous experiment, the higher the $\tan\phi$ value, the higher the range in the hammer estimates. The same effect was observed with in the previous experiment, but in reference to the sounding object hardness estimates.

Discussion

For both sets of stimuli an increment in e was associated with an increment of the hardness estimates. The absence of significant effects of frequency on hammer hardness ratings is consistent with results reported by Freed on real sounds [82]. This result points toward the independence of recognition of hammer geometry–independent properties from the geometrical properties of the sounding object. Results collected with the variable $\tan\phi$ set of stimuli point toward a dependence of the recognition of the geometry–independent features of the hammer on the geometry–independent features of the resonator.

5.3.4 Conclusions

Independently of the fact that subjects were asked to estimate the hardness of the sounding object, in the first experiment, or of the hammer, in the second experiment, the effects of the $\tan\phi$ and of the e variables appeared undifferentiated. In both cases, in fact, hardness estimates increased for increasing $\tan\phi$ and e values. This is surprising, given that subjects were informed

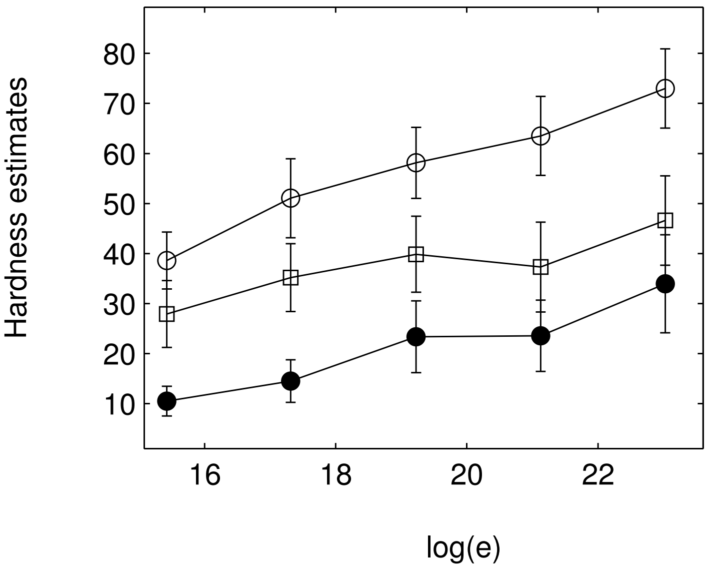


Figure 5.4: Average hammer hardness estimates on a 1-100 scale (1 = very soft 100 = very hard) for the variable $\tan\phi$ set. Filled circles: t_1; empty squares: t_2; empty circles: t_3.

that both the features of the sounding object and of the hammer were varied in the stimuli, that the two tasks were executed by different groups of subjects, and that all of them were given real sounds examples of hardness variations in both the sounding object and in the hammer. This fact would then indicate an inability of subjects to distinguish geometry–independent features of the hammer from geometry–independent features of the sounding object. The frequency of the signal was found to have different effects, depending on whether listeners were asked to estimates the hardness of the hammer or of the sounding object. Hardness estimates of the hammer were found independent from F, which was assumed to model the geometrical features of the sounding object. In contrast F led to significant changes in the estimates of the hardness of the sounding object, in a non–monotonic fashion. This difference reveals the existence of different acoustical criteria for the recognition of the features of the hammer and of the sounding object. However acoustical measurements on the stimuli are necessary to extensively explain subjects responses.

All these results are not consistent with the assumptions of the orthodox ecological approach to sound source recognition. Perturbation variables were found to lead to significant variations in the recognition of the investigated sound source features. Again the validity of these results could be objected because collected with synthetic stimuli. The substantial coincidence of results gathered with this model and results collected by Freed [82] with real sounds, however, suggests the possibility to replicate all these findings with real sounds.

5.4 Overall discussion

Several experiments investigated recognition of two geometry–independent features of the sound source: material of the sounding object and hardness of the hammer and of the sounding object. Given that material represents variations in different mechanical properties, such as hardness, we expected to find similar effects for the recognition of both these features. As expected, both sets of experiments revealed that recognition of these geometry–independent features of the sounding object were influenced by its geometrical properties. Hammer properties influenced hardness scaling but not material categorization. This was explained by the fact that the range of variation in the hammer properties was lower in the first than in the second of these experiments.

Hardness scaling studies revealed the inability of subjects to distinguish variations of the geometry–independent features of the hammer from variations in the geometry–independent features of the sounding object. Even though acoustical analyses for the hardness scaling experiments were not performed, data collected with real sounds by Freed [82] provide a basis to explain, in part, this inability. In this study one of the acoustical parameters that influenced hammer hardness estimates was the slope of the spectral level over time function. This acoustical parameter is strongly related to the T_e parameter used in our material categorization study, as both measure the decay velocity of signal amplitude. Amplitude decay velocity, then, was found to account for material categorization of the sounding object and for hardness estimates of the hammer. Given that material categorization and hardness scaling of the sounding object are related each other, it follows that the overlap of the hardness estimates of the sounding object and of the hammer is partially explained by the fact that both rely on the same acoustical index.

As pointed out above, the only difference between the criteria used to recognize the hardness of the sounding object and the hardness of the hammer stands in the different effects of the geometrical properties of the sounding object, as modelled by the F variable. In fact, this perturbation variable was found to strongly influence the first type of recognition but not the second one. Thus, even though geometry–independent features of the sounding object are confused with geometry–independent features of the hammer, the criteria upon which their recognition is based differ in respect to the influence of the geometrical features of the sounding object.

In all experiments we investigated the recognition of the features of the objects, using a constant type of interaction between them, impact. All the observed effects of the perturbation variables point toward a revision of the veridicality and independence assumptions of the orthodox ecological approach to sound source recognition, in reference to how we recover the features of objects, rather than those of the interaction between objects.

Finally in both sets of experiments, the correspondences between results gathered with real sounds and results gathered with synthetic stimuli support research in sound source recognition conducted with synthetic stimuli. These two approaches likely highlight different aspects of the investigated phenomena. Research carried with real sounds allows direct investigation of the relationship between the physical features of the sound source and recognition, while research carried with synthetic stimuli allow only indirect investigation of this relationship. On the contrary the first type of researches does not allow precise control of the acoustical features, while research conducted on synthetic sounds does. For this reason this latter type of researches makes investigation of the relationship between the acoustical level and sound source recognition straightforward.

The observed misalignments between the physical sound source features, being them real or modelled, and the recognized ones, claim for a general definition of the criterion that we should follow to ascertain the validity of a synthesis models. What we propose is that perceptually effective synthesis models should be able to reproduce the same biases observed in investigations conducted with real sounds.

Chapter 6

Size, shape, and material properties of sound models

Davide Rocchesso, Laura Ottaviani and Federico Fontana
Università di Verona – Department of Computer Science
Verona, Italy
davide.rocchesso@univr.it, ottaviani@sci.univr.it, fontana@sci.univr.it

Federico Avanzini
Università di Padova – Department of Information Engineering
Padova, Italy
avanzini@dei.unipd.it

Recent psychoacoustic studies dealt with the perception of some physical object features, such as shape, size or material. What is in many cases clear from everyday experience, that humans are sensitive to these features, has been proved experimentally in some controlled conditions, using either real-world or synthetic sounds.

A 3D object can be described by its shape, size, material, position and orientation in space, color, surface texture, etc.. Leaving on a side those features that are purely visual (e.g., color), and those features that are only relevant for the "where" auditory subsystem, we are left with shape, size, and material as relevant "ecological" dimensions of sounding objects. This chapter addresses each of these three dimensions as they are representable in signal- or physics-based models. Section 6.1 uses geometry-driven signal models to investigate how size and shape attributes are conveyed to the listener. We bear in mind that ecological acoustic signals tell us a lot about how the objects interact with each other, i.e., about the excitation mechanisms. Indeed, it is the case that different kinds of excitation highlight different features of resonating objects. Even simple impact sounds can elicit the perception of two objects simultaneously, a phenomenon called "phenomenal scission" by some perceptionists (see chapter 2). Section 6.2 stresses the importance of accurate simulation of interaction mechanisms, in particular for calibrating the physical parameters in a sound model in order to render an impression of material by listening to sound sources.

6.1 Spatial features

Of the three object attributes, size, shape, and material, at least the first two are perceptually related in a complex way. The relationship between the ecological properties size and shape reflects a relationship between the perceived sound properties pitch and timbre.

When looking at prior studies in the perception of shape of resonating objects, we find works in:

1D - Carello et al. [49] showed that listeners are able to reliably evaluate, without any particular training, the lengths of rods dropped on a hard surface. Rods are essentially one-dimensional physical systems, so it is a degenerate case of shape perception. Indeed, subjects are perceiving the pitch of complex, non-harmonic tones;

2D - Lakatos et al. [145] showed that the hearing system can estimate the rectangular cross section of struck bars, and Kunkler-Peck and Turvey [142] did similar experiments using suspended rectangular plates. Even though these objects are 3D, attention is focused on their 2D aspect, and it is argued that the distribution of modal frequencies gives cues for shape perception.

Apparently, there is lack of research results in auditory perception of 3D shape. Some direct observations (see chapters 2 and 5) indicate that impact sounds of small solid objects give no clue on the shape of the objects themselves. On the other hand, some prior research [207] indicated that a rough sensation of shape can be elicited by the filtering effect of 3D cavities excited by a source. In that work, the ball-within-a-box (BaBo) model [204] was extended to provide a unified 3D resonator model, based on a feedback delay network, that allows independent control of wall absorption, diffusion, size, and shape. Namely, the shape control is exerted by changing the parameters of allpass filters that are cascaded with the delay lines. In this way, it is possible to have a single computational structure that behaves like a rectangular box, or a sphere, or like an intermediate shape. The availability of such a model raised new questions about the perceptual significance of this shape control.

To investigate the perception of 3D resonator shapes in an experimental framework, we used models of spheres and cubes that are controllable in size and material of the enclosure. We chose to construct impulse responses by additive synthesis, since closed-form expressions of the resonance distributions of cubic and spherical enclosures are available.

A rectangular resonator has a frequency response that is the superposition of harmonic combs, each having a fundamental frequency

$$f_{0,l,m,n} = \frac{c}{2}\sqrt{(l/X)^2 + (m/Y)^2 + (n/Z)^2} \quad , \tag{6.1}$$

where c is the speed of sound, l, m, n is a triple of positive integers with no common divisor, and X, Y, Z are the edge lengths of the box [181].

A spherical resonator has a frequency response that is the superposition of inharmonic combs, each having peaks at the extremal points of spherical Bessel functions. Namely, said z_{ns} the s^{th}

root of the derivative of the n^{th} Bessel function, the resonance frequencies are found at

$$f_{ns} = \frac{c}{2\pi a} z_{ns} \quad , \tag{6.2}$$

where a is the radius of the sphere [178].

The impulse response of a sphere or a cube can be modeled by damping the modes according to the absorption properties of the cavity, introducing some randomization in mode amplitudes in order to simulate different excitation and pickup points, and stopping the quadratic frequency-dependent increase in modal density at the point where the single modes are no longer discriminable. In particular, the decay time of each modal frequency was computed using the Sabine reverberation formula

$$T = 0.163 \frac{V}{\alpha A} \quad , \tag{6.3}$$

where V is volume, A is surface area, and α is the absorption coefficient. The absorption curve was computed by interpolation between the following values, which can be considered as representative of a smooth wood-like enclosure:

$$f = [0, 125, 250, 500, 1000, 2000, 4000, F_s/2] \text{ Hz} \quad ; \tag{6.4}$$

$$\alpha = [0.19, 0.15, 0.11, 0.10, 0.07, 0.06, 0.07, 1.00] \quad , \tag{6.5}$$

and the sample rate was set to $F_s = 22050$ Hz.

This sound model has been used to produce stimuli for tests on perception of cubic and spherical shape and size. In particular, we investigated sizes ranging from 30 cm to 100 cm in diameter. The use of the Sabine formula might be criticized, especially for the range of sizes that we investigated. Indeed, using the Eyring formula or even exact computation of decay time does not make much difference for these values of surface absorption [130]. Moreover, it has been assessed that we are not very sensitive to variations in decay time [234], so we decided to use the simplest formula. This choice, together with the absorption coefficients that we chose, give quite a rich and long impulse response, even too much for a realistic wooden enclosure. However, for the purpose of this experiment it is definitely better to have rich responses so the ear has more chances to discover shape-related information.

6.1.1 Size

When comparing the impulse response of a spherical cavity with that of a cubic cavity, we may notice that one sounds higher than the other. Therefore, there is a pitch relationship between the two shapes. In order to use shape as a control parameter for resonators, for instance in auditory display design, it is important to decouple it from pitch control.

In an experiment [225], we submitted couples of impulse responses to a number of subjects, in random order and random sequence: one of a sphere with fixed diameter and the other of a cube. We used a fixed sphere and thirteen cubes, of which the central one had the same volume as the comparison sphere, while the others had edge length that varied in small steps, equal to Δl,

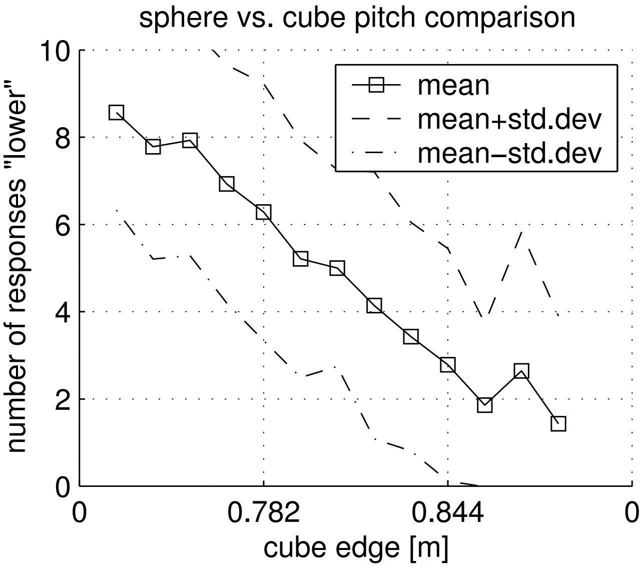


Figure 6.1: Mean and standard deviation of pitch comparison between a sphere ($d = 100$ cm) and cubes differing by an integral number of length JNDs.

calculated by converting frequency Just Noticeable Differences (JND), as found in psychoacoustic textbooks and measured for pure tones, into length differences:

$$\Delta l = \frac{c}{2}\left(\frac{1}{f_0 - \Delta f}\right) - l_0 \quad , \tag{6.6}$$

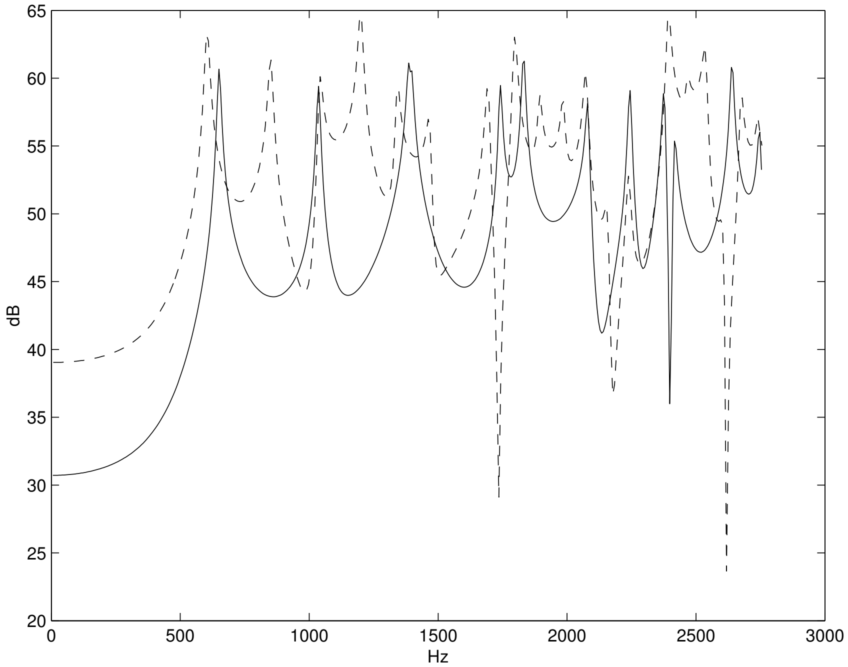
where c is the speed of sound in the cavity, and $l_0 = c/(2f_0)$ is the reference size length. For instance, for a reference size length $l_0 = 1$ m, the length JND is about 18 mm.

Therefore, we had a central cube with the same volume as the fixed comparison sphere, 6 cubes smaller than the central one, and 6 bigger.

Each subject was asked to listen to all the sphere-cube pairs, each repeated ten times. The whole set of 130 couples was played in random order. The question was: "Is the second sound higher or lower in pitch than the first sound?".

The experiment was repeated for two values of sphere diameter, $d = 36$ cm and $d = 100$ cm. Results for the 100 cm sphere and for 14 subjects are plotted in Figure 6.1. A large part of the deviation from the mean curve is due to two outliers who could segregate more than one pitch in the cube impulse response. These two subjects, indeed skilled musicians, seemed to be using a analytic listening mode rather than the holistic listening mode used by the other subjects [30]. This is symptomatic of how the same sonic signal can carry different information for different kinds of listeners.

In both experiments, the Point of Subjective Equality (i.e., where the mean curve crosses the 50% horizontal line) is found where the two shapes are roughly equalized in volume, thus meaning that a good pitch (or size) equalization is obtained with equal volumes. This is an important finding



in an ecological perspective. However, the next question is how to give a “reasonable” value (in Hz) to this
perceived pitch. Conventional spectral analysis does not give any hint to find a

Figure 6.2 plots the frequency responses of a sphere of diameter $d = 36$ cm and a cube with the same volume. Clearly, the pitch of the two cavities in this shape comparison task can not be trivially associated with the fundamental frequency.

In a second experiment [225], we tried to “measure” the perceived pitch comparing the spherical impulse response with an exponentially-damped sinusoid. We chose the spherical response because it was perceived by the subjects to have a higher “pitch strength” than that of the cube. In this pitch comparison test, we used a simple staircase method [152], keeping the ball impulse response fixed (diameter $d = 0.5$ m), and varying the frequency of the test sine wave. The participants to the experiment were 28, that are 25 computer science students, two of the authors, and a research collaborator. They listened to a couple of sounds two times before answering to the question: “Is the first sound higher in pitch than the second sound?”. With the staircase method each subject could converge to a well-defined value (in Hz) of the pitch of the sphere.

Figure 6.3 reports the histogram that shows the distribution of frequency values. It can be noticed that one fourth of the subjects estimated the perceived pitch in the range $[460, 480]$ Hz. Within this range, the estimated pitches were 444.8, 455.8, 461.9, 468.1, 468.9, 473.3, 474.2 Hz. The lowest partial of the spherical frequency response is found at 463.8 Hz. Thus, one fourth of the

Figure 6.3: Distribution of the pitch estimation values of a ball ($d = 0.5$ m).

subjects found a pitch close to the lowest partial. Moreover, Figure 6.3 shows that seven subjects estimated the pitch in the range $[640, 760]$ Hz, that is a neighborhood of the second partial (found at 742 Hz).

In this latter experiment, the difference in the nature of the two stimuli is such that a different, more analytic, listening mode is triggered. Indeed, many subjects tried to match one of the partials of the reference sound. Interestingly, most subjects found this pitch comparison task difficult, because of the different identity of the two stimuli.

On the other hand, from the first experiment it seems that, when a subject tries to match the pitch of two different shapes, he or she is able to extract some information about the physical volume of the two shapes and to match those volumes. That experiment provides evidence of direct perception of volume of cavities, even though the question asked to the listeners did not mention volumes at all.

As far as the use of sound of shapes in information sonification is concerned, one should be aware of the problems related to the relationship size/shape, or pitch/timbre, and to the listening attitude of different listeners, especially if pitch is used as a sonification parameter.

6.1.2 Shape

Besides the size perception of different shapes, it is interesting to understand whether a cubic or a spherical resonator can be correctly classified. If this is the case, a scatter plot where data are represented by simple geometric objects may be sonified by direct shape sonification.

Since most people never experienced the impulse response of small cavities, we made some informal tests by convolving “natural” sounds with the cavity impulse response, to see what could be a suitable excitation signal. The constraint to be remembered is the need of exciting a large part of the frequency response without destroying the identity of the source. We tried with an anechoic voice source, but results were poor, probably because the voice has a strong harmonic content and only a few resonances of the frequency response can be excited in a short time segment. A source such as an applauding audience turned out to be unsuitable because its identity changes dramatically when it is filtered by a one-meter box. Therefore we were looking for a sound source that keeps its identity and that is rich enough to reveal the resonances of the cavities.

We chose a snare drum pattern as a source to generate the stimuli set consisting of 10 sounds, i.e. five couples of stimuli, one from a sphere and one from a cube of the same volume, each couple different from the others for their volumes. The diameters of the spheres were 100 cm, 90 cm, 70 cm, 50 cm, and 30 cm.

The 19 volunteers, who participated to the experiment, listened to isolated sounds, belonging to the stimuli set and each one repeated 10 times in random order. Therefore, the whole experiment consisted in listening to 100 sounds. For each of them, the participants had to say whether it was produced in a spherical or in a cubic enclosure.

The experiment was preceded by a training phase, when the subjects could listen to training sounds as many times as they liked, and they could read the shape that each sound came from. For generating the training sounds, we chose spheres having different sizes than the ones used in the main experiment (diameters 106 cm, 60 cm, and 36 cm), because we wanted to avoid memory effects and to assess the generalization ability of the subjects. In principle, a good listener should be able to decouple shape from pitch during training and to apply the cues of shape perception to other pitches.

One might argue that shape recognition should be assessed without any training. However, as it was pointed out in [207], the auditory shape recognition task is difficult for most subjects just because in real life we can use other senses to get shape information more reliably. This might not be the case for blind subjects [172] but with our (sighted) subjects, we found that training was necessary. Therefore, the task may be described as classification by matching.

The experiment [225] showed that the average listener is able to classify sounds played in different shapes with an accuracy larger than 60% when the cavity diameter is, at least, equal to 50 cm. This is not a very strong result. Again, a major reason is that there were some notable outliers. In particular, some subjects classified resonators of certain sizes (especially the smaller ones) consistently with the same label. This may be due to a non-accurate training or to a mental association between pitch and shape. On the other hand, there were subjects who performed very well.

One of them is visually impaired and her responses are depicted in Figure 6.4. It is clear that for this subject the task was easy for larger volumes and more difficult for smaller volumes. This specific case indicates that there are good possibilities for direct shape sonification, at least for certain subjects.

An auditory feature related to shape is brightness, which is often measured by the spectral centroid and which plays a key role in sound source recognition [171]. We noticed that, when subjects were asked to give a qualitative description of the sounds played in the two different

Figure 6.4: Results of the shape classification for the best subject. Color black in the lower row (and white in the top row) indicates that spheres (or cubes) have been consistently classified throughout the experiment. Grey-levels are used to indicate mixed responses.

cavity shapes, they often said that the sphere sounded brighter. Therefore, we decided to investigate this aspect. Moreover, in order to study the sound characteristics involved in shape recognition, we analyzed the spectral patterns, by means of auditory models. We report the results of these investigations, that are interesting to examine further.

For measuring the brightness of the impulse responses we used a routine that computes the spectral centroid of sounds [205]. For cavities with wooden walls and volume of 1 m^3 the centroid is located at 5570 Hz and 5760 Hz for the cube and the sphere, respectively. This change in brightness must be expected since, for equal volume, the sphere has a smaller surface area than the cube. Therefore, absorption is greater for the cube than the sphere.

The effect of brightness was smoothed, as listeners had to listen to random combinations of pitch (size) and shape; furthermore, brightness is pitch-dependent as well.

We examined the spectral patterns of the spherical and cubic enclosures by analyzing the correlogram calculated from their impulse responses. The correlogram is a representation of sound as a function of time, frequency, and periodicity [222]: Each sound frame is passed through a cochlear model, then split into a number of cochlear channels, each one representing a certain frequency band. Cochlear channels are nonlinearly spaced, according to the critical band model. The signal

Figure 6.5: Correlograms for the cube (edge 0.5 m) and for the sphere having the same volume.

in each critical band is self-correlated to highlight its periodicities. The autocorrelation magnitude can be expressed in gray levels in a 2D plot.

Since the cavities are linear and time invariant, a single frame gives enough information to characterize the behavior of the whole impulse response. We used the impulse responses of resonators having marble walls, just because the images are more contrasted. However, the use of wooden resonators would not have changed the visual appearance of the correlogram patterns appreciably. Figure 6.5 depicts the correlogram of a cube (sized 0.5 m) and a sphere having the same volume. If we superimpose them, we notice the existence of some patterns that, we conjecture, can be read as the *signature* of a particular shape.

It is interesting to notice that the cube has more than one vertical alignment of peaks, this suggesting that expert listeners would be in principle enabled to hear more than one pitch. Moreover, we observed that the curved pattern of the sphere becomes more evident as long as size is increased. Conversely, for small spheres it is barely noticeable: this is confirmed by corresponding fair results in the discrimination of shape for small cavities.

Even though we should be cautious in concluding that these results reflect the kind of pre-processing that is used by our hearing system to discriminate between shapes, nevertheless the correlogram proved to be a useful tool for detecting shapes from sound.

Our analysis also suggests that only few resonances are necessary to form the patterns seen in Figure 6.5. It is likely that no more resonances must be properly located, if we want to convey a sense of shape.

Continuous rendering of morphed shapes

Here we present a method for *sonifying morphing shapes*. By this method we can test whether listeners perceive the *roundness* of morphing shapes, i.e., whether specific cues exist that characterize shape morphing, and whether such cues are conveyed by cartoon models of morphing resonators.

We devised a method for sonifying objects having intermediate shapes between the sphere and the cube, then realizing a model that runs in real time on inexpensive hardware. This model can be used for rendering morphing shapes in 3D virtual scenarios.

The model must be informed about the shapes to render. A specific set of shapes was selected that could be driven by a direct and meaningful morphing parameter, in a way that the control layer relied on a lightweight mapping strategy.

We chose *superquadrics* [15] to form a set of geometries that could be mapped using a few parameters. In particular, a sub-family of superquadrics described by the equation

$$|x|^\gamma + |y|^\gamma + |z|^\gamma = 1 \tag{6.7}$$

(it can be seen that they represent a set of ellipsoidal geometries) fitted our investigation. In fact, changes in shape are simply determined by varying γ, that acts for that family like a morphing parameter:

- sphere: $\gamma = 2$;
- ellipsoid between sphere and cube: $2 < \gamma < \infty$;
- cube: $\gamma \to \infty$.

To implement this 3D resonator model we used a bank of second-order bandpass (peaking) filters, each filter being tuned to a prominent peak in the magnitude response of the superellipsoid. Such a filter bank satisfies the requirement of simplicity and efficiency, and allows "sound morphing" by smoothly interpolating between superellipsoids of different geometry. This is done, for each specific geometry, by mapping γ onto the corresponding filter coefficients, holding condition $2 < \gamma < \infty$.

Since the modal positions are known by theory only in the limit cases of the cube and the sphere, we first calculated the responses of several superellipsoids off-line, with enough precision [78]; then we approximated those responses using filter banks properly tuned. Morphing between two adjacent responses was obtained by linearly interpolating the filter coefficients

Figure 6.6: Trajectories for the lowest resonance frequencies of a superellipsoid morphed from a sphere to cube.

between the values producing the frequency peaks in the two cases, which were previously calculated off-line.

Similarly, different acquisition points in a given resonator were reproduced by smoothly switching on and off the resonance peaks (again calculated off-line) that were present in one acquisition point and not in the adjacent one; otherwise, when one peak was present (absent) in both acquisition points then the corresponding filter was kept active (off) during the whole migration of the acquisition point. Though, we finally decided to pay less attention to the peak amplitudes as they in practice depend on several factors, such as the excitation and acquisition point positions and the internal reflection properties of the resonator.

For what we have said, the resonances thus follow trajectories obtained by connecting neighbor peak positions. These trajectories occasionally split up and rejoin (Figure 6.6).

Moreover, for reasons of symmetry some resonant frequencies can disappear or be strongly reduced while moving the excitation/acquisition points toward the center of the resonator. We accordingly lowered the related peak amplitudes, or cleared them off.

It must be noticed that the frequency trajectories cannot accurately describe the morphing process; they are rather to be taken as a simple (yet perceptually meaningful) approximation, from a number of “sample” shapes, of the continuously changing magnitude response of a morphing object.

During the morphing process the volume of the superellipsoids was kept constant, so respect-ing the volume-pitch relation [208]. This volume constancy leads to a slight decrease of the fundamental mode position when moving toward the cube. Simultaneously, the overall perceived brightness (related to the frequency centroid of the spectrum) decreases when using a fixed number of filters having constant amplitudes.

Several strategies could be conjectured for compensating for this change in brightness. Though, we did not reputed this correction to be necessary, since the effect does not seem to be strong enough to mask other sound features related to the shape morphing.

6.2 Material

Rendering an impression of object material is not always possible or cost effective in graphics. On the other hand, physics-based sound models give the possibility to embed material properties with almost no computational overhead.

Wildes and Richards [253] developed theoretical considerations about the relationship between the damping (equivalently, decay) characteristics of a vibrating object and the auditory perception of its material. Two recent studies provided some experimental basis to the conjectures formulated in [253], but results were not in accordance. On the one hand, Lutfi and Oh [163] found that changes in the decay time are not easily perceived by listeners, while changes in the fundamental frequency seem to be a more salient cue. On the other hand, Klatzky et al. [137] showed that decay plays a much larger role than pitch in affecting judgment, and therefore confirmed predictions by Wildes and Richards. Both of these studies used synthetic stimuli generated with additive synthesis of damped sinusoids.

A physically-based approach was taken by Djoharian [65], who developed viscoelastic models in the context of modal synthesis and showed that finite difference models of resonators can be "dressed" with a specific material quality. Indeed, the physical parameters of the models can be chosen to fit a given frequency-damping characteristic, which is taken as the sound signature of the material. Sound examples provided by Djoharian convinced many researchers of the importance and effectiveness of materials in sound communication.

Any sound practitioner is aware of the importance of the attack in percussive sound timbres. Nonetheless, none of the above mentioned works made use of impact models, instead using simple impulse responses with no attack transients. It remains to be proved to what extent is material perception affected when realistic and complex impacts are used.

Using a method for artifact-free simulation of non-linear dynamic systems [24], we developed a digital impact model that simulates collision between two modal resonators. This digital hammer is based on a well known model in impact mechanics [167], and is described in detail in chapter 8. We produced synthetic auditory stimuli with this model, and used the stimuli for investigating material perception through listening tests. In order to keep the number of model parameters low, the modal resonator in the synthesis algorithm was parametrized so to have only one mode.

Two acoustic parameters were chosen for controlling synthesis of stimuli, namely pitch f_0 and quality factor q_o. Using our resonator model, this latter parameter relates to decay characteristics

Figure 6.7: Proportion of subjects who recognized a certain material for each stimulus.

via the equation $q_o = \pi f_o t_e$, where t_e is the time for the sound to decay by a factor $1/e$. We synthesized 100 stimuli using five equally log-spaced pitches from 1000 to 2000 Hz and 20 equally log-spaced quality factors from 5 to 5000; these extremal q_o values correspond to typical values found in rubber and aluminum, respectively. In a recent study on plucked string sounds, Tolonen and Järveläinen [234] found that relatively large deviations (between -25% and $+40\%$) in decay time are not perceived by listeners. Accordingly with the values we chose, the relative lower/upper spacings between q_o values are $-31\%/+44\%$.

In the listening test, subjects were asked to listen to these 100 sounds and to indicate the material of the sound sources, choosing among a set of four material classes: rubber, wood, glass and steel. Each sound was played only once. The subjects were 22 volunteers, both expert and non-expert listeners, all reporting normal hearing.

Figure 6.7 summarizes test results: it shows the proportion of subjects responses for each material, as a function of the two acoustic cues (pitch and quality factor). The inter-subject agreements (proximity of the response proportions to 0 or 1) are qualitatively consistent with indications given by [253], namely (1) q_o tends to be the most significant cue and (2) q_o is in increasing order for rubber, wood, glass and steel. A slight dependence on pitch can be noticed: rubber and glass tend to be preferred at high pitches, while wood and steel are more often chosen at low pitches. It appears clearly that the upper and lower halves of the q_0 range are well separated, while materials within each of these two regions are less easily discriminated. In particular, it is evident from the figure that the regions corresponding to glass and steel are largely overlapping, while ranges for rubber and wood are better delimited. The attribution of material qualities to sound is more problematic when recordings are used. Chapter 5 extensively covers this case.

Chapter 7

Experiments on gestures: walking, running, and hitting

Roberto Bresin and Sofia Dahl
Kungl Tekniska Högskolan – Department of Speech, Music, and Hearing
Stockholm, Sweden
roberto.bresin@speech.kth.se, sofia.dahl@speech.kth.se

7.1 Introduction

In the recent past, sound synthesis techniques have achieved remarkable results in reproducing real sounds, like those of musical instruments. Unfortunately most of these techniques focus only on the "perfect" synthesis of isolated sounds. For example, the concatenation of these synthesized sounds in computer-controlled expressive performances often leads to unpleasant effects and artifacts. In order to overcome these problems Dannenberg and Derenyi [60] proposed a performance system that generates functions for the control of instruments, which is based on spectral interpolation synthesis.

Sound control can be more straightforward if sounds are generated with physics-based techniques that give access to control parameters directly connected to sound source characteristics, as size, elasticity, mass and shape. In this way sound models that respond to physical gestures can be developed. This is exactly what happens in music performance when the player acts with her/his body on the mechanics of the instrument, thus changing its acoustical behavior. It is therefore interesting to look at music performance research in order to identify the relevant gestures in sound control. In the following sections outcomes from studies on the analysis of music performance are considered.

7.2 Control models

The relations between music performance and body motion have been investigated in the recent past. Musicians use their body in a variety of ways to produce sound. Pianists use shoulders, arms, hands, and feet; trumpet players make great use of their lungs and lips; singers put into actions their glottis, breathing system, phonatory system and use expressive body postures to render their interpretation. Percussion players, generally use large body movements to be able to hit the right spot at the right time. Dahl [58] recently studied movements and timing of percussionists when playing an interleaved accent in drumming. The movement analysis showed that drummers prepared for the accented stroke by raising the drumstick up to a greater height, thus arriving at the striking point with greater velocity. In another study drummers showed a tendency to prefer auditory feedback to tactile feedback [59]. These and other observations of percussionists' behavior have being considered for the modeling of a control model for physics-based sound models of percussion instruments. This control model could be extended to the control of impact sound models where the human action is used to manipulate the sound source.

The research on music performance at KTH, conducted over a period of almost three decades, has resulted in about thirty so-called performance rules. These rules, implemented in a program called Director Musices[1] [84], allow reproduction and simulation of different aspects of the expressive rendering of a music score. It has been demonstrated that rules can be combined and set up in such a way that emotionally different renderings of the same piece of music can be obtained [33]. The results from experiments with emotion rendering showed that in music performance, emotional coloring corresponds to an enhancement of the musical structure. A parallel can be drawn with hyper- and hypo-articulation in speech; the quality and quantity of vowels and consonants vary with the speaker's emotional state or the intended emotional communication [156]. Yet, the structure of phrases and the meaning of the speech remain unchanged. In particular, the rendering of emotions in both music and speech can be achieved, and recognized, by controlling only a few acoustic cues [33, 131]. This is done in a stereotype and/or "cartoonified" way that finds its visual correspondent in email emoticons. Therefore cartoon sounds can be produced not only by simplifying physics-based models, but also by controlling their parameters in appropriate ways.

In the following sections it is discussed how studies in music performance can be useful when designing a control model for footstep sounds. Analysis of gesture and timing of percussionists are also presented; outcomes can be implemented in the control model of impact sound models.

7.3 Walking and running

As a first sound control case we chose that of locomotion sounds. In particular we considered walking and running sounds. In a previous study Li and coworkers [153] demonstrated the human ability to perceive source characteristics of a natural auditory event. They ask subjects to classify

[1] Director Musices http://www.speech.kth.se/music/performance/download/

Figure 7.1: Spectrogram of a walking sound on concrete floor (16 footsteps).

Figure 7.2: Spectrogram of walking sound on gravel (16 footsteps).

the gender of a human walker. Subjects could correctly classify walking sounds as being produced by men or women. Subjects showed also ability in identifying the gender in walking sequences even with both male and female walkers wearing the same male's shoes.

Sounds In their study Li and coworkers considered walking sounds on a hardwood stage in a art theater. From various analyses applied on the walking sounds, they found a relationship between auditory events and acoustic structure. In particular male walking sounds were characterized by "(1) low spectral mean and mode (2) high values of skewness, kurtosis, and low-frequency slope, and (3) low to small high-frequency energy". On the other hand female walking sounds were characterized by "(1) high spectral mean and mode, and significant high spectral energy".

In the present study we considered sounds of walking and running footstep sequences produced by a male subject running on gravel. The choice was motivated by the assumption that (1) an isolated footstep sound produced by a walker on a hard surface would be perceived as unnatural, i.e. mechanical, (2) the sound of an isolated footstep on gravel would still sound natural because of its more noisy and rich spectrum (Figures 7.1, 7.2, 7.3 and 7.4).

Figure 7.3: Long time average spectrogram (LTAS) of a walking sound on gravel (40 footsteps).

Figure 7.4: Long time average spectrum (LTAS) of a running sound on gravel (60 footsteps).

Tempo When walking, a double support phase is created when both feet are on the ground at the same time, thus there is a step overlap time. This is shown also in the spectrogram of three footsteps of Figure 7.5; there is no silent interval between to adjacent steps. This phenomenon is similar to legato articulation. Figure 7.6 plots the key-overlap time (KOT) and the double support phase duration (Tdsu) as a function of the inter-onset interval (IOI) and of half of the stride cycle duration (Tc/2), respectively. The great inter-subject variation in both walking and legato playing, along with bio-mechanical differences, made quantitative matching impossible. Nevertheless, the tendency to overlap is clearly common to piano playing and walking. Also common is the increase of the overlap with increasing IOI and increasing (Tc/2), respectively.

Both jumping and running contain a flight phase, during which neither foot has contact with the ground. This has also a visual representation in the spectrogram of three footsteps of Figure 7.7; there is a clear silent interval between two adjacent steps. This is somewhat similar to staccato articulation in piano performance. In Figure 7.7 the flight time (Tair), and key-detach time (KDT) are plotted as a function of half of stride cycle duration (Tc/2) and of IOI. The plots for Tair correspond to typical step frequency in running. The plots for KDT represent mezzostaccato and staccato performed with different expressive intentions as reported by Bresin and Battel [32]. The

Figure 7.5: Spectrogram of three steps extracted from the walking sound on gravel of Figure 7.1.

similarities suggest that it would be worthwhile to explore the perception of legato and staccato in formal listening experiments

7.3.1 Controlling the sounds of walking and running

Among the performance rules developed at KTH there are rules acting on a short timescale (micro-level rules), and rules acting on a long timescale (macro-level rules) [83]. Examples of the first class of rules include the "Score Legato Articulation" rule, which realizes the acoustical overlap between adjacent notes marked legato in the score, and the "Score Staccato Articulation" rule, which renders notes marked staccato in the score [31]. A macro-level rule is the "Final Ritard" rule that realizes the final ritardando typical in Baroque music [86]. Relations between these three rules and body motion have been found. In particular Friberg and Sundberg demonstrated how their model of final ritardando was derived from measurements of stopping runners, and in the previous paragraph we pointed out analogies in timing between walking and legato, running and staccato. In both cases human locomotion is related to time control in music performance.

Friberg et al. [87] recently investigated the common association of music with motion in a direct way. Measurements of the ground reaction force by the foot during different gaits were transferred to sound by using the vertical force curve as sound level envelopes for tones played at different tempi. The results from the three listening tests were consistent and indicated that each tone (corresponding to a particular gait) could clearly be categorized in terms of motion.

These analogies between locomotion and music performance open to a challenging field for the design of new control models for artificial walking sound patterns, and in general for sound control models based on locomotion. In particular a model for humanized walking and one for stopping runners were implemented in two `pd` patches. Both patches control the timing of the sound of one step on gravel.

The control model for humanized walking was used for controlling the timing of the sound of one step of a person walking on gravel. As for the automatic expressive performance of a music score, two performance rules were used to control the timing of a sequence of steps. The

Figure 7.6: The double support phase (Tdsu, filled symbols) and the key overlap time (KOT, open symbols) plotted as function of half of stride cycle duration (Tc/2) and of IOI. The plots for Tdsu correspond to walking at step frequency as reported by Nilsson and Thorstensson [185, 186]. The KOT curves are the same as in Figure 7.1, reproducing data reported by Repp [201], Bresin and Battel [32], MacKenzie and Van Eerd [165].

two rules were the "Score Legato Articulation" rule and the "Phrase Arch" rule. The first rule, as mentioned above, presents similarities with walking. The "Phrase Arch" rule is used in music performance for the rendering of accelerandi and rallentandi. This rule is modeled according to velocity changes in hand movements between two fixed points on a plane [74]. When it was used in listening tests of automatic music performances, the time changes caused by applying this rule were classified as "resembling human gestures" [131]. The "Phrase Arch" rule was then considered to be interesting for use in controlling tempo changes in walking patterns and combined with the "Score Legato Articulation" rule. In Figure 7.9 the tempo curve and the spectrogram for walking sounds, produced by the model, is presented.

The control model for stopping runners was implemented with a direct application of the "Final Ritard" rule to the control of tempo changes on the sound of running on gravel.

In the following we describe a pilot experiment where the control models presented here are tested for the first time.

7.3.2 Pilot experiment: listening to walking and running sounds

A listening test comparing step sound sequences without control, and sequences rendered by the control models presented here, was conducted. We wanted to find out if (1) listeners could discriminate between walking and running sounds in general and (2) if they were able to correctly

Figure 7.7: Spectrogram of three steps extracted from the running sound on gravel of Figure 7.2.

classify the different types of motion produced by the control models.

Stimuli Eight sound stimuli were used. They were 4 walking sounds and 4 running sounds. The walking sounds were the following; (1) a sequence of footsteps of a man walking on gravel indicated with W_{SEQ} in the following, (2) one footstep sound extracted from stimuli W_{SEQ} (W_{1STEP}), (3) a sequence of footsteps obtained by looping the same footstep sound (W_{NOM}), (4) a sequence of footsteps obtained applying the “Phrase arch” and the “Score legato articulation” rules (W_{LP}). The running sounds were; (1) a sequence of footsteps of a man running on gravel (R_{SEQ}); (2) one footstep sound extracted from stimuli R_{SEQ} (R_{1STEP}), (3) a sequence of footsteps obtained by looping the same footstep sound (R_{PD}), obtained with a `pd` patch, (4) a sequence of footsteps obtained applying the “Final Ritard” rule (R_{RIT}).

Subjects and procedure The subjects were four, 2 females and 2 males. Their average age was 28. The subjects all worked at the Speech Music Hearing Department of KTH, Stockholm.

Subjects listened to the examples individually over Sennheiser HD433 adjusted to a comfortable level. Each subject was instructed to identify for each example (1) if it was a walking or running sound and (2) if the sound was human or mechanical. The responses were automatically recorded by means of the Visor software system, specially designed for listening tests [105]. Listeners could listen as many times as needed to the each sound stimuli.

Results and discussion We conducted a preliminary statistical analysis of the results. In Figure 7.10 average subjects’ choices are plotted. The Y axis represents the scale from Walking, value 0, to Running, value 1000, with 500 corresponding to a neutral choice. It emerges that all walking and running sounds were correctly classified as walking and running respectively. This including the footstep sequences generated with the control models proposed above. This means that listeners have in average no problem to recognize this kind of stimuli. It is however interesting to notice how the stimuli corresponding to a single footstep were classified with less precision than the other sounds in the same class (walking or running). In particular one of the subjects classified

Figure 7.8: The time when both feet are in the air (Tair, filled symbols) and the key detached time (KDT, open symbols) plotted as function of half of stride cycle duration (Tc/2) and of IOI. The plots for Tair correspond to normal frequency steps in running [185, 186]. The KDT for mezzostaccato (KDR = 25%) is defined in the Oxford Concise Dictionary of Music [135]. The values for the other KDTs are reported in works by Bresin and Battel [32] and Bresin and Widmer [34].

the stimulus W_{1STEP} as a running sound. The R_{1STEP} was classified as mechanical, although it is a real sound. This could be the reason why the sequences of footstep sounds produced by the control models were all classified as mechanical, since these sequences loop the same footstep.

7.3.3 Discussion

In this section we proposed a model for controlling the production of footstep sound. This control model is rule-based and it is derived from analogies between music performance and body motion that have been observed in previous works of the KTH research group. Even though this is still an exploratory work, a listening test was conducted. In this pilot experiment, the model was applied to the control of walking and running sampled sounds. The main result was that subjects correctly classified different types of motion produced by the model.

Recently, the sampled sounds were substituted by a physics-based model of crumpling sounds [77]. This model allows the modeling of expressive walking and running patterns by varying various foot-related parameters, such as the pressure on the ground, and the size of the foot. A model of expressive walking could be useful for characterizing footsteps of avatars in virtual reality applications.

Figure 7.9: Tempo curve and overlap time curve used for controlling a walking sound.

The proposed rule-based approach for sound control is the first step toward the design of more general control models that respond to physical gestures. Continuing on the same research direction, other models were developed based on rule-based control of parameters in sound models. In the following sections and in chapter 11 studies on the expressive gestures of percussionists and disk jockeys (Dj) are discussed. From these studies two rule-based control models were derived, one for percussive instruments and one for Dj scratching. These models can produce a natural expressive variation of the control parameters of sound models in accordance with the dynamics

Figure 7.10: Mean classification values, with pooled subjects, for the scale Running (1000) - Walking (0). W_{SEQ} is a sequence of footsteps of a man walking on gravel; W_{1STEP} is one footstep sound extracted from stimuli W_{SEQ}; W_{NOM} is a sequence of footsteps obtained by looping the same footstep sound; W_{LP} is a sequence of footsteps obtained applying the "Phrase arch" and the "Score legato articulation" rules; R_{SEQ} is a sequence of footsteps of a man running on gravel; R_{1STEP} is one footstep sound extracted from stimuli R_{SEQ}; R_{PD} is a sequence of footsteps obtained by looping the same footstep sound obtained with a `pd` patch; R_{RIT} is a sequence of footsteps obtained applying the "Final Ritard" rule.

of the gestures.

7.4 Modeling gestures of percussionists: the preparation of a stroke

Playing the drums. Mastering rhythm and tempo requires a playing technique, which can be adapted to the feedback from the instrument. Percussion playing, in general, can require that the player perform the same rhythm on several different instruments with different physical properties (e.g. the stiffness of the drumhead and the mass, hardness and shape of the mallet). Therefore it seems reasonable to assume that a skilled player, when possible, will take advantage of these different properties to minimize the experienced effort.

Four percussion players' strategies for performing an accented stroke were studied by capturing movement trajectories. The main objective was to investigate what kind of movement strategies the players used when playing interleaved accented strokes, the hypothesis being that accented strokes would be initiated from a greater height than the unaccented strokes. Other questions addressed were whether there would be any differences in movement patterns or striking velocities depending on playing conditions; dynamic level, tempo, or striking surface.

7.4.1 Method

Three professional percussionists and one amateur played on a force plate with markers on the drumstick, hand, and lower and upper arm. The movements were recorded using a motion detection system (selspot), tracking the displacement of the markers at 400 Hz. The rhythmic pattern - an ostinato with interleaved accents every fourth stroke - was performed at three dynamic levels (pp, mf, and ff), at three tempi (116, 160, and 200 beats per minute), and on three different striking surfaces added to the force plate (soft, normal, and hard).

The analysis was concentrated on the vertical displacement of the drumstick at the initiation of a stroke, the *preparatory height*, and the velocity right before impact, the *striking velocity*. Both these measures were extracted from the vertical displacement of the marker on the drumstick.

The *metric location* of a stroke, that is, the position of the stroke in the measure was of special interest. The mean values were therefore calculated and presented with respect to the metric location so that a data value referring to metric location #1 is averaged across all first strokes in the measure for the specified player and playing condition. For instance, Figure 7.12 show the average striking velocities for the four players playing on the normal surface, where metric locations #1, #2, and #3 are unaccented strokes and #4 the accented stroke.

7.4.2 Results

Movement trajectories. The analysis showed that the four players used movement strategies to play the accented strokes. The movements were maintained consistently within each player, but the movement trajectories differed considerably between the players (see Figure 7.11). All subjects raised the stick to a greater height before the accented stroke. In Figure 7, paper II, the average preparatory height for the four subjects are seen. The figure shows how the players increased the preparatory heights with increasing dynamic level and in preparation for the accented stroke (beat #4).

Striking Velocities. The characteristics of the players' individual movement patterns were reflected in the players' striking velocities. The observed preparatory heights corresponded well with the striking velocities. The most influential parameter on the movement patterns was the dynamic level. Comparing the striking surfaces, the players tended to increase striking velocity when playing on the soft surface, but decrease striking velocity for the hard surface.

Both the preparatory heights and the striking velocities show that the main difference between the playing styles of the drummers was the emphasis on the accented stroke as compared to the

Figure 7.11: Motion trajectories captured from four markers on the drumstick, and the subjects' hand, lower and upper arm. The panels show the movement patterns of the four subjects S1 (upper left), S2 (upper right), S3 (lower left), and S4 (lower right) as seen from the players' left side. The numbers correspond to the location of the markers as seen in the illustration above. Each panel includes approximately four measures at mf, 200 beats per minute. The preparatory movements for the accented stroke can be seen as a larger loop compared to that of the unaccented strokes. The players' drumstick, hand, lower and upper arm are involved to different extent in the movements, reflecting the different playing styles. Subjects S1 and S3 (left column) are mainly playing in orchestras, while subjects S2 and S4 (right column) mainly play the drumset.

unaccented. For instance, players S1 and S2 produced similar average striking velocities for the unaccented strokes, but while S1 played the accented strokes on average with a velocity 1.5 times higher than for unaccented, the striking velocity for S2's accented stroke was almost five times the unaccented.

Figure 7.13 show a comparison between how player S1 and S2 emphasize the accented stroke compared to the unaccented, for different tempi and dynamic levels. The figure shows a linear plane fitted to the measured striking velocities for all unaccented strokes (stroke #2 (bottom plane), and the accented strokes (stroke #4, top plane) for the two players playing on the normal surface. As illustrated by the inclination of the planes in the figure, tempo and dynamic level has different influence on the two players' emphasis on the accented strokes.

Two of the players (S2 and S4) also showed a slight decrease in striking velocity for stroke #3, the stroke preceding the accent. The explanation can be found in the preparatory movements. Both these players initiate the preparation for the accented stroke by moving the hand and wrist upwards. To reach the height from which the accented stroke is initiated in ample time the hand starts the upward movement already before the preceding hit.

7.4.3 Discussion

The movement patterns of the players were clearly reflected in the striking velocities. It is possible that the differences between the players' movement strategies and emphasis on the accented beat compared to the unaccented beat could refer to the different background of the players. Player S1 and S3 are mainly active in the symphonic and military orchestral traditions, while S2 and S4 mainly play the drumset. In orchestral playing an accent, although emphasized, should not be *over*emphasized. In contrast, many genres using drumset playing often encourages big dynamic differences between stressed and unstressed beats. In fact, unstressed notes are sometimes played as soft as possible; "ghost notes".

In Figure 7.11 trajectories from the analyzed percussionists can be seen, representing the two playing styles. The most emerging feature in the playing style of the two players was the amount of emphasis on the accented note compared to the unaccented. While all subjects raised the drumstick up to a greater height in preparation for the accented stroke, the ratio between the preparatory height for unaccented and accented strokes varied considerably between players. In the figure the symphonic player (S1) lifts the stick only slightly higher for the accented note, the drumset player (S2) on the other hand, displays great differences in the preparations for the two kinds of strokes, (the accented stroke seen as a large loop compared to the unaccented strokes).

The movement trajectories with their differences in preparatory height were reflected in the striking velocity. Figure 7.12 shows the average striking velocity for the four subjects. Each data point is averaged across its position in the measure, its *metric location*. Thus a data point at metric location #1 is averaged across all first strokes in the measure, and so on. All strokes occurring at metric location #4 were accented and it is clearly seen how all players raised the striking velocity for this stroke. The amount of emphasis on the accented note differed between players, but also with dynamic level and tempo.

The influence of different dynamic level and tempi on strokes played by subjects S1's and S2's is illustrated in Figure 7.13. The figure shows linear planes fitted to the striking velocities for an unaccented stroke (stroke #2, bottom plane), and the accented strokes (stroke #4, top plane). As

Figure 7.12: Average striking velocity for the four subjects. The values are displayed versus their metric location, where metric location #4 contains all accented strokes. The different emphasis the four players place on the accented stroke as compared to the unaccented is clearly seen. Most markedly is the differences between the two players S1 and S2.

the inclination of the planes in the figure clearly shows, tempo and dynamic level has different influence on the two players' emphasis on the accented strokes. The fit of the linear plane with data is not optimal, but was used as a simple estimate on the players different emphasis on the accented stroke.

7.5 Auditory feedback vs. tactile feedback in drumming

Percussion players are able to control their own movements with great precision under different conditions. In doing this they use their own movement strategies, but also a variety of expressive gestures (which also are able to convey expression to observers).

use a variety of expressive gestures when striking their instrument and they Therefore, an analysis of gestures as performed by percussionists can provide useful insights for the design of control models for impact sound models.

Players strive to acquire a playing technique that will allow them to play as long, fast and loud as required with sufficient precision. Precisely how long, fast and loud a player needs to play is dependent on the music and the performing conditions. Contemporary music that uses drum-set is usually performed together with electronically amplified instruments in larger music halls.

Figure 7.13: Striking velocities for players S1 (top panel) and S2 (bottom panel) playing at all tempi and dynamic levels on the normal surface. The graphs show a linear plane fitted to the measured striking velocities for an unaccented stroke (stroke #2, bottom plane), and the accented strokes (stroke #4, top plane). The fit of a linear plane to data is not optimal, but was used as an estimation of the drummers' different interpretation of the accented stroke compared to the unaccented.

This calls for a higher range of dynamic level than does the symphonic orchestra, which instead demands great precision at soft dynamic levels.

Regardless of their musical genre, skilled players seem to have developed a playing technique that is flexible yet efficient. A study of four percussionists with different backgrounds performing single strokes with interleaved accents [58] have shown that each of the four subjects displayed a characteristic movement pattern that was maintained very consistently.

The rebound. Like many instrumentalists percussionists are very dependent on the tactile feedback from the instrument. The time that the drum stick is in contact with the drum head is very short (usually between 5-10 ms [57]) but determines the timbre and quality of the stroke. The time is so short that the player has small opportunities to adjust anything once the drumstick hits the drum head. To get the desired result the movement has to be right throughout the whole stroke, from beginning to end, including the final position after the rebound.

Contradicting feedback. When a player is performing on electronic drum pads and synthesized drums the relationship between the tactile and the auditory feedback is somewhat different than for the acoustical drum. The recorded or synthesized sounds make it possible to change the acoustical properties of the drum very easily, without affecting the physical properties of the instrument. A change of sound is no longer necessarily connected to a corresponding change in the tactile feedback. Consequently the sounding stroke may not correspond exactly to the characteristics of the played stroke. For the player, the mismatch between sound and tactile feedback introduces a conflict. This feature would not be possible when playing acoustical drums, and hence playing electronic drum pads is to conquer a different instrument.

A physical model of an instrument supplies a sound source that responds naturally to gestures during playing. Such a model, however, can introduce some delay in the response. Electronic instruments dependent on electronic amplification may well be subject to delays through signal processing or even because of too large distances between loudspeakers and players. It is therefore important to know to what extent such perturbations can be tolerated.

Delayed Auditory Feedback - DAF. As new electronically instruments make conflicts and contradictions between different feedbacks possible, it is an interesting question which sensory feedback a player will adjust to. Manipulating the auditory feedback by introducing a delay is easily done and would also have a strong relation to the normal playing situation where delays of the auditory signals can occur.

Earlier studies have investigated the effect of Delayed Auditory Feedback (DAF) in music performance. Gates [91] used delay values in the range 100 ms to 1.05 s for keyboardists. They found that players tended to slow down in tempo, most markedly at a delay of 270 ms. Finney [72] reported large errors in performance for pianists subjected to DAF during playing. In his study the delayed feedback caused more discrepancies in interhand coordination and introduced more errors compared to the conditions with combined delay and pitch modulation of the feedback to the player. Pfordresher and Palmer [194] found that the spread in timing decreased for certain delay values that coincided with subdivisions of the performed tempo.

In these investigations the delay values studied have been well above the just noticeable differences for tempo perturbation. In studies of time discrimination in isochronous sequences the just noticeable difference ranged between 1 and 9% of the inter-onset interval (IOI), depending on the type of perturbation (see [85] for a survey). It could therefore also be interesting to study the effects on musical performance of smaller delay values.

7.5.1 Experiments

To investigate the impairment of delays in auditory feedback on rhythm production, we studied drumming on electronic percussion instruments with delayed auditory feedback. The investigation has compared players trying to maintain a steady tempo without an external time keeper, to players synchronizing with a metronome. The hypothesis was that if a player has to synchronize with another audio source, i.e. other musicians, he/she will try to do this by matching sound with sound for as long as possible. There should, however, come a certain point when the time delay is so large that the player no longer can disregard the discrepancies. The player will then have to make an active choice of which feedback to rely on and this should cause a change in the temporal errors produced. For the drummers performing without metronome we expected a steeper shift in tempo for about the same amount of delay.

Pilot experiments

In two pilot experiments the effect of delayed auditory feedback on tempo synchronization was investigated. The pilot experiments are described in detail in [59]. In a first investigation the Max Mathews radio-baton [25] was used as a percussion instrument. The output files were analyzed with respect to the time difference between the onset of the baton stroke and the metronome. In a second pilot experiment the radio baton was exchanged for a commercial drum pad (Clavia ddrum [61]) and the spread in IOI were studied.

The two pilot experiments supported the assumption that the player indeed tries to match sound with sound when playing with a metronome. In Figure 7.14 it's clearly seen how the player compensates for the delay by initiating the strokes earlier, striving to match the delayed auditory feedback with the sound of the metronome. The second pilot experiment also indicated that a possible break-point could be sought between 40-55 ms, after which the timing error seemed to increase.

Main experiment

Method and subjects. The experimental set-up used a patch in `pd` [197] to

- control the generation of a MIDI note with the desired auditory delay, and reproducing a percussive sound
- store timing data of the hits
- produce a metronome to indicate the tempo. Subjects listened through a pair of closed headphones that blocked the direct audio feedback from the playing.

Through the patch the experimenter was able to control the tempo of the metronome and the delay of the auditory feedback to the player in real time.

10 drummers participated as subjects. The subjects were instructed to play single strokes with their preferred hand and to maintain the tempo indicated by the metronome. After adjusting the

Figure 7.14: Time difference between radio-baton stick hits and metronome versus introduced delay for one of the subjects in the pilot experiment. With increased delay the player strives to place the stroke earlier and earlier to match the sound with the sound of the metronome.

stool and position of the ddrum pad to a comfortable height and checking the sound level in the head phones, the experiment started. The player was seated so that there was no direct view of the experimenter managing the manipulating PC.

Each subject performed at two different tempi, 120 and 92 beats per minute (BPM) The nominal beat separation for the two tempi are 500 and 652 ms respectively. For each recording session the player started at one of the two tempi (randomly chosen), which was indicated by the metronome through the headphones. If the subject was playing without a metronome the metronome was turned off when the player had adjusted to the tempo (after approximately 10 strokes).

The experimenter introduced the delay in steps from 0 to Δt and then back to 0. Each step in Δt was maintained for about 10 to 17 strokes, so that the player was not prepared for a tempo change. After each step in Δt there would be a period (of about the same length) with $\Delta t = 0$, before introducing a new Δt. An example of how the changes in auditory feedback were introduced to the player is shown in Figure 7.15.

At each tempo the first session, *upward* session, the first perturbation started with a Δt of 5 ms. Δt was then increased by 5 ms each time the delay returned until about 160 ms, or until the player failed to continue playing. In the following session, *downward* session, the delay started at a value little above where the subjects stopped playing in the previous session and Δt was reduced with 5 ms for each occurrence until the zero-level was reached. After the first two sessions the player had a recess for about 5 minutes before the two remaining sessions at the other tempo.

Figure 7.15: Example of raw data from one of the subjects performing with metronome. The top curve shows the successive IOIs encircling the nominal tempo, 92 BPM (nominal beat separation 652 ms) indicated by the dashed line. At the bottom of the figure the amount of delay and the onset times can be seen. It can be noted that the larger changes between adjacent IOIs do not necessarily coincide with the shifts in Δt.

Analysis The analysis was concentrated to the spread in IOIs. The last nine IOIs produced before each change in Δt were pooled together and the spread of data was calculated as the standard deviation. To make comparisons between different tempi possible the IOIs were also normalized to the average IOI across the same last 9 intervals for each each Δt. These normalized IOIs were then, in turn, pooled together and the spread of data calculated. For $\Delta t = 0$ only the first 9 intervals *before* the perturbations began were included. The range in Δts covered varied between sessions. For the range up to 145 ms were all subjects represented, although some subjects played up to a delay of 200 ms.

A first visual inspection of data revealed that one of the subjects playing with metronome performed very poorly. As this subject had stated that drums were not his main instrument it seemed likely that the many and large disturbances in data had an origin in poor control of the drumstick, and this subject was therefore omitted from the remaining analysis.

There were some gaps in the data set. The gaps were due to unintended ‘double-clicks’ from the experimenter while controlling the patch, which caused the introduction of a new Δt before the intended 10 strokes had been played. In all, 80 out of a total of 9720 strokes (9 intervals x 30 delay values x 4 sessions x 9 subjects) were missing, in the range up to $\Delta t = 145$. For some values of Δt almost all strokes would be played but, more common, there might also be no strokes recorded for a certain Δt during for a specific session. The missing strokes occurred only for one delay value for the subjects performing with metronome ($\Delta t = 20$ ms) but for 9 different delay values for the subjects performing without metronome ($\Delta t = 5, 10, 15, 20, 35, 40, 60$ and 105 ms).

The gaps in data resulted in a weaker foundation for the pooling and the calculation of the standard deviation for the concerned Δts. However, as the tests were tiring, aborting and restarting a test session was not considered to benefit the experiment. Subjects would probably have been more likely to produce errors due to lack of concentration than due to conflicting sensory feedback.

7.5.2 Results

General impressions during the experiment were that players at first easily coped with the delay. In the beginning of the first upward session there were but minor disturbances in timing and playing appeared easy. As the auditory delay was increased subjects tended to increase the striking force to achieve more tactile feedback from the ddrum pad. Drummers performing without metronome generally decreased tempo throughout the session, but some subjects (both performing with and without metronome) also tended to either increase or decrease tempo during the periods where the delay was applied.

Figure 7.16 show examples of one subject playing with metronome and another playing without. The figure shows the mean IOI played for each delay value (Δt), and the vertical error bars indicate the standard deviation. For the subject performing with metronome (top panels) there is a section when the player clearly fails to maintain the tempo. In the upward session (top left) the difficulties are seen as considerable deviations for delay values between 75 and 110 ms. As delay is increased further, however, the player once more manages to follow the metronome. For the downward session (top right) there is a corresponding interval with increased variability between 135 and 90 ms, after which the average tempo settles down to metronome tempo.

For the subject performing without metronome there are clear difficulties to maintain the original tempo already from start. For each delay value there are few large discrepancies between successive strokes. The standard deviations are fairly constant throughout the whole session, and never come close to the large deviations displayed in the top panels. A more telling indication of the players' difficulties is the drift in tempo. For both the upward and downward session the rate at which the player decreases tempo is slowest for delay values less than 80 ms. In the upward session, however, the player suddenly starts to increase the average tempo for delay values over 170 ms. No increase in tempo can be seen in the downward session, but the full span of the tempo drift can be seen in the jump between delay value $\Delta t = 5$ and $\Delta t = 0$. These values used in the averages for $\Delta t = 0$ were always taken from the strokes in the beginning of the session, before the perturbations, and are thus not connected to the preceding data points in the same way.

Comparing the performances for all subjects some general observations can be made:

1. Large differences between adjacent IOIs did not necessarily coincide with the the times where the delay was applied or removed. Rather, it would normally take about 4-5 strokes after the shift before the player would 'loose track' in the rhythm.

Figure 7.16: Two subjects playing with (top) and without metronome (bottom). The figures show the mean IOI (averaged across the last 9 strokes for each delay value, Δt), with the standard deviation. Test sessions with increasing delay (Δts from 5 ms and up) are seen to the left, while right panels show the sessions with decreasing delay (Δts from 180 ms and down). The tempo indicated by the metronome (500 ms = 120 BPM) is represented by a dashed line.

2. Large ‘dips’ in the average IOI accompanied by large standard deviations (as seen in top panels in Figure 7.16) are mainly due to strokes at approximately the double tempo. Although errors, they indicate that the subjects tried to maintain tempo by subdivision. There were, however, also plenty of strokes that were not close to any subdivision.

3. A certain training effect can be discerned for the subjects playing with metronome. For the first upward session there are more severe disturbances and hence larger spreads compared to the remaining sessions.

4. Comparing the overall performances between subjects performing with and without metro-

nome across the delay range 0-145 ms, the standard deviations were generally higher for subjects playing with metronome (mean value 5.2% compared to 3.1% for subjects playing without metronome). In this group also the maximal standard deviations appeared, reaching 14%.

5. For the subjects performing without metronome there were more difficulties to maintain the slower tempo (92 beats per minute). For the higher tempo (120 beats per minute) there were equal numbers of increases as decreases in tempo.

7.5.3 Discussion

To conclude the findings from the experiments there are clear effects of the DAF on the performances of the drummers also for delay values below the previously explored ranges. Difficulties to maintain tempo and large errors in the timing of individual strokes appeared more frequently for certain ranges in delay values. Possible breakpoints could be sought in two ranges; one between 20-55 ms, and another between 80 and 130 ms.

The large shifts between adjacent intervals and the larger variations in standard deviations for the players performing with metronome compared to those performing without, could have several sources. The results seem to point toward two options:

1. Since the metronome continues to mark the original tempo, it makes coping with the effects from the delayed auditory feedback more ‘urgent’ for these players. Players performing without metronome, on the other hand, were able to drift further in tempo without adjusting.

2. As pointed out by Madison [166], playing along with a metronome is in itself not a trivial task.

That large errors and tempo drifts occur already for values well below 100 ms makes it clear that there is a conflict between the perceived audio signal and the tactile feedback. It is interesting though, to see how the players manages to discard the audio signal and return to ‘normal’ standard deviations once the delay is large enough. This indicates that tactile feedback in drumming is something that can be relied on once the delayed auditive feedback has exceeded a certain ‘critical’ value. It was also very evident how the subjects increased striking force for delay values that disturbed them. In the downward sessions some players also made almost abrupt changes in striking force as the delay values once more passed under the lower breakpoint. The effect with the increased striking force could be further explored in additional investigations.

Chapter 8

Low-level models: resonators, interactions, surface textures

Federico Avanzini
Università di Padova – Department of Information Engineering
Padova, Italy
avanzini@dei.unipd.it

Matthias Rath, Davide Rocchesso and Laura Ottaviani
Università di Verona – Department of Computer Science
Verona, Italy
rath@sci.univr.it, davide.rocchesso@univr.it, ottaviani@sci.univr.it

8.1 Introduction

"Traditional" techniques of sound synthesis (e.g. additive, subtractive, FM) are based on, and accessed through, signal theoretic parameters and tools. While deep research has established a close connection to conventional musical terms, such as pitch, timbre or loudness, newer psychoacoustic works [92] point out that the nature of everyday listening is rather different. From the *ecological* viewpoint, auditory perception delivers information about a listener's surrounding, i.e. objects in this surrounding and their interactions, mostly without awareness of and beyond attributes of musical listening. The common use of wavetables, i.e. the playback of prerecorded sound files, can probably be seen as the standard reaction to the severe restrictions existing in former methods of sound generation. That approach is still signal-based, and not satisfactory in many contexts, such as in human-computer interaction or in interactive virtual environments, due to the static nature of the produced sounds.

In contrast, we use the term *acoustic modeling* to refer to the development of "sound objects" that incorporate a (possibly) complex responsive acoustic behavior, expressive in the sense of ecological hearing, rather than fixed isolated signals. Physically-based models offer a viable way to synthesize naturally behaving sounds from computational structures that can easily interact with the environment and respond to physical input parameters. Various modeling approaches can be used: Van den Doel et al. [238, 237] proposed modal synthesis [2] as an efficient yet accurate framework for describing the acoustic properties of objects. Contact forces are used to drive the modal synthesizer, under the assumption that the sound-producing phenomena are linear, thus being representable as source - filter systems. For non-interactive applications, it has been proposed to generate sound as a side effect of nonlinear finite element simulations [188]. In this way, sounds arising from complex nonlinear phenomena can be simulated, but the heavy computational load prevents the use of the method in interactive applications. Physical models are widely developed in the computer music community, especially using the waveguide simulation paradigm [223], but their main application has been the faithful simulation of existing musical instruments.

Although real sounds hereby serve as an orientation, realistic simulation is not necessarily the perfect goal: simplifications which preserve and possibly exaggerate certain acoustic aspects, while losing others considered less important, are often preferred. Besides being more effective in conveying certain information, such "cartoonifications" are often cheaper to implement, just like graphical icons are both clearer and easier to draw than photo-realistic pictures. Can the idea of audio cartoons suggest an approach to sound design, that fills the gap between simulation or arrangement of concrete sounds and abstract musical expression?

The design approach outlined in this chapter can be roughly referred to as *low-level* modeling. The basic physical mechanisms involved in sound generation are accounted for, including the description of resonating structures, as well as various interaction modalities, such as impact and continuous contact (friction). Section 8.2 describes an efficient structure for describing resonating objects, based on the modal analysis/synthesis approach. Section 8.3 presents models for the nonlinear interaction forces which arise during collision and friction between two modal resonators, and describes an efficient model for rendering texture of surfaces with variable degree of roughness. Section 8.4 discusses the implementation of these contact models as real-time modules, and addresses the problem of control. Finally, details about the synthesis algorithms are given in the appendix 8.A. This is especially meant to help the understanding of the numerical techniques used for developing real-time algorithms, and provides detailed equations and pseudo-code. Chapter 9 will make use of these sound models for developing *high-level* strategies for sound presentation.

8.2 Modal resonators

8.2.1 Continuous-time model

The simplest possible representation of a mechanical oscillating system is a second-order linear oscillator of the form

$$\ddot{x}^{(r)}(t) + g^{(r)}\dot{x}^{(r)}(t) + \left[\omega^{(r)}\right]^2 x^{(r)}(t) = \frac{1}{m^{(r)}} f_{ext}(t) \quad , \tag{8.1}$$

where $x^{(r)}$ is the oscillator displacement and f_{ext} represents any external driving force, while the parameters $\omega^{(r)}$ and $g^{(r)}$ are the oscillator center frequency and damping coefficient, respectively. The parameter $1/m^{(r)}$ controls the "inertial" properties of the oscillator (note that $m^{(r)}$ has the dimension of a mass). Such a one-dimensional model provides a basic description of the resonator in terms of its pitch $\omega^{(r)}$ and quality factor $q^{(r)} = \omega^{(r)}/g^{(r)}$. The parameter $g^{(r)}$ relates to the decay properties of the impulse response of system (8.1): specifically, the relation $t_e = 2/g^{(r)}$ holds, where t_e is the $1/e$ decay time of the impulse response.

However, in most cases a single-mode oscillator is not enough to produce interesting and spectrally-rich sounds. A slightly more sophisticated model is obtained by parallel connection of N oscillators such as that of Eq. (8.1). By choosing a different center frequency $\omega_l^{(r)}$ $(l = 1 \dots N)$ for each oscillator, it is possible to account for a set $\{\omega_l^{(r)}\}_{l=1}^N$ of partials of the resonator spectrum. A set of N decoupled modal resonators excited by the same external force can be described by means of a multi-variable generalization of Eq. (8.1). In matrix form, this can be written as

$$\begin{bmatrix} \ddot{x}_1^{(r)}(t) \\ \vdots \\ \ddot{x}_N^{(r)}(t) \end{bmatrix} + \boldsymbol{G}^{(r)} \begin{bmatrix} \dot{x}_1^{(r)}(t) \\ \vdots \\ \dot{x}_N^{(r)}(t) \end{bmatrix} + \left[\boldsymbol{\Omega}^{(r)}\right]^2 \begin{bmatrix} x_1^{(r)}(t) \\ \vdots \\ x_N^{(r)}(t) \end{bmatrix} = \boldsymbol{m}^{(r)} f_{ext}(t) \quad , \tag{8.2}$$

where the matrices are given by

$$\boldsymbol{\Omega}^{(r)} = \begin{bmatrix} \omega_1^{(r)} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \omega_N^{(r)} \end{bmatrix}, \quad \boldsymbol{G}^{(r)} = \begin{bmatrix} g_1^{(r)} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & g_N^{(r)} \end{bmatrix}, \quad \boldsymbol{m}^{(r)} = \begin{bmatrix} 1/m_1^{(r)} \\ \vdots \\ 1/m_N^{(r)} \end{bmatrix} \quad . \tag{8.3}$$

When a distributed resonating object is modeled as a chain of N masses connected with springs and dampers, the resulting system is composed of N coupled equations [180]. However, the theory of modal analysis [2] shows that it is generally possible to find a transformation matrix $\boldsymbol{T} = \{t_{jl}\}_{j,l=1}^N$ which diagonalizes the system and turns it into a set of decoupled equations. The transformed variables $\{x_l^{(r)}\}_{l=1}^N$ are generally referred to as *modal displacements*. The displacement x_j and velocity v_j of the resonating object at a given point $j = 1 \dots N$ are then given by

$$x_j = \sum_{l=1}^{N} t_{jl} x_l^{(r)} \quad \text{and} \quad \dot{x}_j = \sum_{l=1}^{N} t_{jl} \dot{x}_l^{(r)} \quad . \tag{8.4}$$

The modal description given by Eqs. (8.2), (8.4) provides a high degree of controllability. The damping coefficients $g_l^{(r)}$ control the decay times of each exponentially-decaying mode of the resonator. The frequencies $\omega_l^{(r)}$ can be chosen to reproduce spectra corresponding to various geometries of 1D, 2D and 3D resonators. As an example, the first N resonances of a cavity can be mapped into the modal frequencies of the N oscillators, and morphing between different shapes can be obtained by designing appropriate trajectories for each of these resonances.

(1,1) (1,2)

Figure 8.1: A circular membrane displaced from its rest position according to the spatial shape of mode(1,1) (left) and mode(1,2) (right). The frequencies of vibration along these axes are 1.593 and 2.917 times that of mode(0,1) (the "fundamental").

In the remainder of this chapter, the quantities $m_l^{(r)}$ are referred to as *modal masses*, while the quantities $1/m_l^{(r)}$ are referred to as *modal weights*. Note that by allowing the modal masses to vary for each oscillator, the matrix $\boldsymbol{m}^{(r)}$ can be generalized to give control on the amounts of energy provided to each oscillator (see section 8.2.2 below). This permits simulation of position-dependent interaction, in that different interaction points excite the resonator modes in different ways.

8.2.2 Position-dependent excitation

Figure 8.1 shows a membrane which is displaced from its rest position in such a way that only one single mode is set into vibration. The distance of each point of the membrane from the "rest plane" is proportional to the weighting factor $1/m^{(r)}$ of the mode at this position. Note that the intersections of the mode–shape with the rest plane (i.e., the *nodal lines*) remain fixed during the entire cycle of the modal vibration. Therefore, the modal weights at these positions are 0 (equivalently, the modal masses tend to infinity). Correspondingly, an external force applied at these node lines does not excite the mode at all.

In order for the resonator model (8.2) to account for such a situation, the weights $1/m_l^{(r)}$ must be made position-dependent. In other words, the $(N \times 1)$ matrix $\boldsymbol{m}^{(r)}$ must be generalized by defining a $(N \times N)$ matrix $\boldsymbol{M}^{(r)}$, whose element (l, j) is the modal weight of mode l at interaction point j.

There are several possible approaches to gain the position dependent weights. In the case of a finite one dimensional system of point masses with linear interaction forces, modal parameters are exactly found through standard matrix calculations. Most systems of interest of course do not fit these assumptions. In some cases the differential equations of distributed systems can be solved analytically, giving the modal parameters; this holds for several symmetric problems such as circular or rectangular membranes.

Alternatively, either accurate numerical simulations (e.g. wave-guide mesh methods) or "real" physical measurements can be used. Impulse responses computed (or recorded) at various interaction points then form a basis for the extraction of modal parameters. The acoustic "robustness" of the modal description allows convincing approximations on the basis of microphone-recorded signals of e.g. an object struck at different points, despite all the involved inaccuracies: spatially distributed interaction, as well as wave distribution in air, provide signals that are quite far from impulse/frequency responses at single points.

Qualitative observations on modal shapes can be effectively used in a context of cartoonification: for modes of higher frequencies the number of nodal lines increases and their spatial distance decreases accordingly. One consequence is that for higher modes even small inaccuracies in interaction or pickup position may result in strongly different modal weights, so that an element of randomization can here add "naturalness". In the case of vibrating strings, membranes, or clamped bars, the boundary is a nodal line for all the modes, and the higher modes gradually gain importance over the lower modes as the interaction point is shifted toward the boundary. This phenomenon can be well noticed for a drum: if the membrane is struck close to the rim, the excited sound becomes "sharper", as the energy distribution in the frequency spectrum is shifted upward ("rimshots"). For a clamped bar higher partials are dominant near the fixed end, whereas lower frequencies are stronger for strokes close to the free vibrating boundary (noticeable in sound adjustments of electromechanical pianos).

Similar considerations apply to points of symmetry: some resonant modes, those with modal shapes antisymmetric to central axes, are not excited when the driving force is applied at the center of a round or square membrane. They consequently disappear "bottom–up" when approaching the center point. For 3D resonators, such as cavities, one can generally say that the most effective modal excitation is obtained at the boundary. For instance, in a rectangular room seven modes over eight have a nodal plane passing through the room center, while all the modes are excited at a corner.

8.2.3 Discrete-time equations

The continuous-time system (8.3) is discretized using the bilinear transformation[1], which is usually interpreted as a s-to-z mapping between the Laplace and the Z domains:

$$s = 2F_s \frac{1 - z^{-1}}{1 + z^{-1}} \quad . \tag{8.5}$$

The bilinear transformation is one appealing discretization technique for various reasons. First, its order of accuracy can be seen [146] to be two. Second, the transformation preserves the order of the system (e.g., the second-order equation (8.1) is turned into a second-order difference equation by the bilinear transformation). Finally, the transformation is stable, since the left-half s-plane is mapped by Eq. (8.5) into the unit z-circle. Consequently, the bilinear transformation provides a reasonable trade-off between accuracy and efficiency. On the other hand, some of its properties can be seen as drawbacks in this context. Noticeably, it introduces frequency warping [177], and it is an implicit method (which has some consequences on the resulting numerical equations, as discussed in appendix 8.A.1 below).

After applying transformation (8.5) to system (8.2), the resulting discrete-time system appears as a parallel filter bank of second-order low-pass resonant filters, each one accounting for one specific mode of the resonator. The output of the filter bank can be taken to be the weighted sum (8.4) of either the modal displacement or the modal velocities. Each of the filters can be accessed to its parameters of center-frequency $\omega_l^{(r)}$ and damping coefficient $g_l^{(r)}$ (or, equivalently, decay time t_{el}).

The complete numerical system takes the form

$$\begin{cases} \boldsymbol{x}_l^{(r)}(n) = \boldsymbol{A}_l^{(r)} \boldsymbol{x}_l^{(r)}(n-1) + \boldsymbol{b}_l^{(r)}[y(n) + y(n-1)] \quad , \quad (\text{for} \quad l = 1 \ldots N) \\ \boldsymbol{x}_j(n) = \sum_{l=1}^{N} t_{jl} \boldsymbol{x}_l^{(r)}(n) \quad , \quad (\text{for} \quad j = 1 \ldots N) \\ y(n) = f_{ext}(n) \end{cases} \quad , \tag{8.6}$$

where the vectors $\boldsymbol{x}_l^{(r)}$ and $\boldsymbol{x}_j$ are defined as $\boldsymbol{x}_l^{(r)} = \begin{bmatrix} x_l^{(r)} \\ \dot{x}_l^{(r)} \end{bmatrix}$, and $\boldsymbol{x}_j = \begin{bmatrix} x_j \\ v_j \end{bmatrix}$, respectively. The matrices $\boldsymbol{A}_l^{(r)}$ and $\boldsymbol{b}_l^{(r)}$ are found to be

$$\boldsymbol{A}_l^{(r)} = \frac{1}{\Delta_l^{(r)}} \begin{bmatrix} \Delta_l^{(r)} - \left[\omega_l^{(r)}\right]^2/2 & F_s \\ -F_s \left[\omega_l^{(r)}\right]^2 & 2F_s^2 - \Delta_l^{(r)} \end{bmatrix} \quad , \quad \boldsymbol{b}_l^{(r)} = \frac{1}{m_l^{(r)}} \cdot \frac{1}{4\Delta^{(r)}} \begin{bmatrix} 1 \\ 2F_s \end{bmatrix} \quad , \tag{8.7}$$

where the quantity $\Delta_l^{(r)}$ is given by $\Delta_l^{(r)} = F_s^2 + F_s/t_{el} + \left[\omega_l^{(r)}\right]^2/4$.

[1] Also known in the numerical analysis literature as the one-step *Adams-Moulton method* [146]

Figure 8.2: Cartoon impact between two modal resonators.

Note that the modal weights $1/m_l^{(r)}$ only appear in the definition of $\boldsymbol{b}_l^{(r)}$, which controls the extent to which mode l is excited by the external force f_{ext}. Following the discussion in section 8.2.2 above, multiple excitation points can be modeled by attaching an additional index $j = 1 \dots N$ to the modal masses.

8.3 Interactions

8.3.1 Impact

In this section we construct a continuous-time impact model between two modal resonators, following the “cartoonification” approach that has informed the research activities of the SOb project: Figure 8.2 shows that the sound model is controlled through a small number of parameters, which are clearly related either to the resonating objects or to the interaction force. The precise meaning and role of the parameters depicted in Figure 8.2 will be explained in the remainder of this section.

Impact models have been widely studied in musical acoustics, mainly in relation with hammer-string interaction in the piano. If the contact area between the two colliding objects is assumed to be small (ideally, a point), the simplest model [111] states a polynomial dependence of the contact force f on the hammer felt compression x:

$$f(x(t)) = \begin{cases} k[x(t)]^{\alpha} & x > 0 \\ 0 & x \leq 0 \end{cases} , \tag{8.8}$$

The compression x at the contact point is computed as the difference between hammer and string displacements. Therefore, the condition $x > 0$ states that there is actual felt compression, while the complementary condition $x \leq 0$ says that the two objects are not in contact. The parameter k is the force *stiffness*, and the exponent α depends on the local geometry around the contact area. As an example, in an ideal impact between two spherical object α takes the value 1.5. Typical experimental values in a piano hammer felt range from 1.5 to 3.5, with no definite trend from bass to treble.

More realistic models have to take into account the hysteresis effects involved in the interaction. As an example, it is known that the force-compression characteristic in a piano hammer exhibits a hysteretic behavior, such that loading and unloading of the hammer felt are not alike. In particular, the dynamic force-compression characteristic is strongly dependent on the hammer normal velocity before collision. In order to account for these phenomena, Stulov [230] proposed an improved model where the contact force possesses history-dependent properties. The idea, which is taken from the general theory of mechanics of solids, is that the spring stiffness k in Eq. (8.8) has to be replaced by a time-dependent operator. Consequently, according to Stulov the contact force can be modeled as

$$f(x(t),t) = \begin{cases} k[1 - h_r(t)] * [x(t)^\alpha] & x > 0 \\ 0 & x \leq 0 \end{cases} , \tag{8.9}$$

where $*$ stands for the continuous-time convolution operator, and $h_r(t) = \frac{\epsilon}{\tau} e^{-t/\tau}$ is a *relaxation function* that controls the "memory" of the material. In fact, by rewriting the convolution explicitly the Stulov force is seen to be:

$$f(x(t),t) = kx(t)^\alpha - \frac{\epsilon}{\tau} e^{-t/\tau} \int_0^t e^{\xi/\tau} x(\xi)^\alpha \, d\xi \quad \text{for} \quad x > 0 \quad . \tag{8.10}$$

The Stulov model has proved to be successful in fitting experimental data where a hammer strikes a massive surface, and force, acceleration, displacement signal are recorded. Borin and De Poli [23] showed that it can be implemented numerically without significant losses in accuracy, stability and efficiency with respect to the simpler model (8.8).

Useful results on impact modeling are also found from studies in robotics. Physical modeling of contact events is indeed a relevant issue in dynamic simulations of robotic systems, when physical contact with the environment is required in order for the system to execute its assigned task (for example, handling of parts by an industrial manipulator during assembly tasks, or manipulator collisions with unknown objects when operating in an unstructured environment). Marhefka and Orin [167] provide a detailed discussion of a collision model that was originally proposed by Hunt and Crossley [122]. Under the hypothesis that the contact surface is small, Hunt and Crossley proposed the following form for the contact force f:

$$f(x(t),v(t)) = \begin{cases} kx(t)^\alpha + \lambda x(t)^\alpha \cdot v(t) = kx(t)^\alpha \left(1 + \mu v(t)\right) & x > 0 \\ 0 & x \leq 0 \end{cases} , \tag{8.11}$$

where $v(t) = \dot{x}(t)$ is the compression velocity, and k and α are defined as above. The parameter λ is the force damping weight, and $\mu = \lambda/k$ is a mathematically convenient term which is called "viscoelastic characteristic" by Marhefka and Orin. Similarly to Eqs. (8.8) and (8.9), the value of the exponent α depends only on the local geometry around the contact surface. Note that the force model (8.11) includes both an elastic component kx^α and a dissipative term $\lambda x^\alpha v$. Moreover, the dissipative term depends on both x and v, and is zero for zero compression.

Marhefka and Orin have studied the following case: an idealized hammer, described as a lumped mass $m^{(h)}$, strikes a surface. The surface mass is assumed to be much greater than $m^{(h)}$, therefore the surface is assumed not to move during the collision. When the two objects collide, the hammer initial conditions are $x^{(h)}(0) = 0$ (hammer position) and $\dot{x}^{(h)}(0) = -v_{in}$ (hammer normal velocity before collision). Since the surface is assumed not to move, the hammer position and velocity relate to the compression and compression velocity through the equalities $x^{(h)}(t) = -x(t)$, $\dot{x}^{(h)}(t) = -v(t)$. The hammer trajectory is therefore described by the differential equation $m^{(h)}\ddot{x}^{(h)} = f(-x^{(h)}, -\dot{x}^{(h)})$. Then it is shown in [167] that

$$\frac{d(\dot{x}^{(h)})}{dx^{(h)}} = \frac{\dot{v}}{v} = \frac{(\Lambda v + K)\,[x]^\alpha}{v} \quad \Rightarrow \quad \int \frac{v\,dv}{(\Lambda v + K)} = \int [x]^\alpha\,dx \quad , \tag{8.12}$$

where two auxiliary parameters $\Lambda = -\lambda/m^{(h)}$ and $K = -k/m^{(h)}$ have been introduced for clarity. The integral in Eq. (8.12) can be computed explicitly and gives

$$x(v) = \left[\left(\frac{\alpha+1}{\Lambda^2}\right)\left(\Lambda(v - v_{in}) - K\log\left|\frac{K + \Lambda v}{K + \Lambda v_{in}}\right|\right)\right]^{\frac{1}{\alpha+1}} \quad . \tag{8.13}$$

Eq. (8.13) provides x as a function of v, and can therefore be exploited for plotting the phase portrait on the (x, v) plane. This is shown in Figure 8.3a.

Another remark by Marhefka and Orin is concerned with "stickiness" properties of the contact force f. From Eq. (8.11), it can be seen that f becomes inward (or sticky) if $v < v_{lim} := -1/\mu$. However, this limit velocity is never exceeded on a trajectory with initial conditions $x = 0$, $v = v_{in}$, as shown in the phase portrait of Figure 8.3a. The upper half of the plot depicts the trajectories of a hammer which strikes the surface with various normal velocities (trajectories are traveled in clockwise direction). Note that the output velocities after collision v_{out} are always smaller in magnitude than the corresponding v_{in}. Moreover, for increasing v_{in} the resulting v_{out} converges to the limit value v_{lim}. The horizontal line $v = v_{lim}$ corresponds to the trajectory where the elastic and dissipative terms cancel, and therefore the hammer travels from right to left with constant velocity. This horizontal line separates two regions of the phase space, and the lower region is never entered by the upper paths. The lower trajectories are entered for an initial compression $x < 0$ and initial *negative* compression velocity $v_{in} < v_{lim}$. If such conditions are imposed, then one of the lower trajectories is traveled from left to right: the hammer bounces back from the surface, while its velocity decreases in magnitude, due to the dissipative term in the force f.

Figure 8.3b shows the compression-force characteristics during collision. Note that the dissipative term $\lambda x^\alpha v$ introduces hysteresis. In this respect the role of the dissipative term, in the Hunt and Crossley model, is very similar to that of the relaxation function in the Stulov model.

(a) (b)

Figure 8.3: Collision of a hammer with a massive surface for various v_{in}'s; (a) phase portrait, (b) compression-force characteristics. Values for the hammer parameters are $m^{(h)} = 10^{-2}$ [Kg], $k = 1.5 \cdot 10^{11}$ [N/m$^{\alpha}$], $\mu = 0.6$ [s/m], $\alpha = 2.8$, $v_{in} = 1 \ldots 4$ [m/s].

The Hunt and Crossley impact model (8.11) can be used as a coupling mechanism between two modal resonators (described in section 8.2). For clarity, the two objects are denoted with the superscripts (h) and (r), which stand for "hammer" and "resonator", respectively. The two objects interact through the impact contact force $f(x, v)$ given in Eq. (8.11). Assuming that the interaction occurs at point $l = 1 \ldots N^{(h)}$ of the hammer and point $m = 1 \ldots N^{(r)}$ of the resonator, the continuous-time equations of the coupled system are given by:

$$\left\{ \begin{array}{l} \ddot{x}_i^{(h)} + g_i^{(h)} \dot{x}_i^{(h)} + \left[\omega_i^{(h)}\right]^2 x_i^{(h)} = \dfrac{1}{m_{il}^{(h)}} (f_e^{(h)} + f) \quad , \quad (i = 1 \ldots N^{(h)}) \\ \ddot{x}_j^{(r)} + g_j^{(r)} \dot{x}_j^{(r)} + \left[\omega_j^{(r)}\right]^2 x_j^{(r)} = \dfrac{1}{m_{jm}^{(r)}} (f_e^{(r)} - f) \quad , \quad (j = 1 \ldots N^{(r)}) \\ x = \sum_{j=1}^{N^{(r)}} t_{mj}^{(r)} x_j^{(r)} - \sum_{i=1}^{N^{(h)}} t_{li}^{(h)} x_i^{(h)} \quad , \quad v = \sum_{j=1}^{N^{(r)}} t_{mj}^{(r)} \dot{x}_j^{(r)} - \sum_{i=1}^{N^{(h)}} t_{li}^{(h)} \dot{x}_i^{(h)} \quad , \\ f(x, v) = \left\{ \begin{array}{ll} k x(t)^{\alpha} + \lambda x(t)^{\alpha} \cdot v(t), & x > 0 \\ 0, & x \leq 0 \end{array} \right. \quad \text{(impact force)} \end{array} \right. \tag{8.14}$$

where $x_i^{(h)}$ and $x_j^{(r)}$ are the modal displacements of the hammer and the resonator, respectively. The terms $f_e^{(h)}, f_e^{(r)}$ represent external forces, while the integers $N^{(h)}$ and $N^{(r)}$ are the number of modes for the hammer and the resonator, respectively. As a special case, one or both the objects can be a "hammer", i.e. an inertial mass described with one mode, zero spring constant and zero internal damping. As another special case, one object can be a "rigid wall", i.e. a modal object

Figure 8.4: Cartoon friction between two modal resonators.

with an ideally infinite mass.

8.3.2 Friction

The continuous-time friction model presented in this section follows the same "cartoon" approach adopted for the impact model: Figure 8.4 shows that the sound model is controlled through a small number of parameters, which are clearly related either to the resonating objects or to the interaction force. The precise meaning and role of the parameters depicted in Figure 8.4 will be explained in the remainder of this section. Given the non-linear nature of friction, interacting structures with few resonances are able to produce complex and rich sonorities. This is an important issue from a computational viewpoint, since efficient models can be developed that provide realistic simulations of contacting objects. It is necessary to bear in mind, however, that when looking for accurate reproduction of friction phenomena $\ll$ *[...] there are many different mechanisms. To construct a general friction model from physical first principles is simply not possible [...]* $\gg$ [189].

The friction models adopted in the literature of physical modeling of bowed string instruments are typically referred to as *kinetic* or *static* models. Although the two definitions may seem contradictory (kinetic vs. static) at a first glance, they actually refer to the same modeling approach: given a fixed bow pressure, the friction force f is assumed to be a function of the relative velocity only (kinetic models), and the dependence is derived under stationary conditions (static models). An example of parametrization of the steady velocity friction force is given by the following equation:

Figure 8.5: The static friction model (8.15), computed with parameters values $\mu_d = 0.197, \mu_s = 0.975, v_s = 0.1, f_N = 0.3$.

$$f(v) = \text{sgn}(v)\left[f_c + (f_s - f_c)e^{-(v/v_s)^2}\right] \quad , \tag{8.15}$$

where f_c, f_s are the Coulomb force and the stiction force, respectively, while v_s is usually referred to as Stribeck velocity. Figure 8.5 provides a plot of this function. Note in particular that static models are able to account for the so-called Stribeck effect, i.e. the dip in the force at low velocities. The stiction force is always higher than the Coulomb force, and the term $e^{-(v/v_s)^2}$ parametrizes the slope of the dip in the friction force as the relative velocity v increases.

In recent years, a new class of friction models has been developed and exploited for automatic control applications, where small displacements and velocities are involved, and friction modeling and compensation is a very important issue. These are usually referred to as *dynamic* models, since the dependence of friction on the relative sliding velocity is modeled using a differential equation. Being more refined, these models are able to account for more subtle phenomena, one of which is presliding behavior, i.e. the gradual increase of the friction force for very small displacement values. Static and dynamic friction models exhibit the same behavior at high or stationary velocities, but dynamic models provide more accurate simulation of transients [7], which is particularly important for realistic sound synthesis. The difference between static and dynamic models is qualitatively similar to what occurs in reed instrument modeling: it has been shown that dynamic models of the single reed mechanism offer superior sound quality and are capable to reproduce various oscillation regimes found in experiments with real instruments [12].

The first step toward dynamic modeling was proposed by Dahl (see [189] for a review), and

Figure 8.6: The bristle interpretation (a) and the LuGre single-state averaged model (b).

was based on the stress-strain curve of classic solid mechanics. This has been later improved by the so-called *LuGre* model[2] which provides a more detailed description of frictional effects [62]. Specifically, friction is interpreted as the result of a multitude of micro-contacts (bristles), as shown in Figure 8.6a. The LuGre model describes this interaction as a single-state system which represents the average bristle behavior (as in Figure 8.6b).

One drawback of the LuGre model is that it exhibits drift for arbitrarily small external forces, which is not physically consistent. This effect has been explained in [66] by observing that LuGre does not allow purely elastic regime for small displacements: therefore, a class of *elasto-plastic* models has been proposed in [66], where the drawbacks of LuGre are overcome. These models have been applied in [117] to haptic rendering applications. An alternative extension of LuGre has been proposed in [231], which incorporates hysteresis with nonlocal memory in the non-linear friction force. The elasto-plastic models are going to be used in the remainder of this section, and consequently demand a more detailed description.

The pair of equations

$$\begin{aligned} \dot{z}(v,z) &= v\left[1-\alpha(z,v)\frac{z}{z_{ss}(v)}\right] \quad , \\ f(z,\dot{z},v,w) &= \sigma_0 z + \sigma_1\dot{z} + \sigma_2 v + \sigma_3 w \quad , \end{aligned} \tag{8.16}$$

summarizes the elasto-plastic modeling approach. The first equation in (8.16) defines the averaged bristle behavior as a first-order system: z and $\dot{z}$ can be interpreted as the mean bristle displacement and velocity, respectively, while v is the relative velocity. The second equation in (8.16) states that the friction force f results from the sum of three components: an elastic term $\sigma_0 z$, an internal dissipation term $\sigma_1\dot{z}$, and a viscosity term $\sigma_2 v$ which appears in lubricated systems.[3] A fourth component $\sigma_3 w$ is added here to equation (8.16), which is not part of the original formulation by Dupont et al. [66]. The term $w(t)$ is a pseudo-random function of time which introduces noise in the force signal, and is therefore related to surface roughness (see also section 8.3.3 below).

[2] The name derives from LUnd and GREnoble, and refers to the two research groups that have developed the model.

[3] As explained in [189], the viscosity term needs not to be linear and may be a more complicated function of the relative velocity.

The auxiliary functions α and z_{ss} can be parametrized in various ways. Here we follow [62] by defining z_{ss} as

$$z_{ss}(v) = \frac{\text{sgn}(v)}{\sigma_0}\left[f_c + (f_s - f_c)e^{-(v/v_s)^2}\right] \quad , \tag{8.17}$$

where f_c, f_s, and v_s are defined as above (see Eq. 8.15), and the subscript ss in z_{ss} stands for "steady-state". As far as α is concerned, we follow [66] by defining it as

$$\alpha(v, z) = \begin{cases} \left.\begin{array}{ll} 0 & |z| < z_{ba} \\ \alpha_m(v,z) & z_{ba} < |z| < z_{ss}(v) \\ 1 & |z| > z_{ss}(v) \end{array}\right\} & \text{if } \text{sgn}(v) = \text{sgn}(z) \\ 0 & \text{if } \text{sgn}(v) \neq \text{sgn}(z) \end{cases} \quad . \tag{8.18}$$

The function $\alpha_m(v, z)$, which describes the transition between elastic and plastic behavior, is parametrized as

$$\alpha_m(v, z) = \frac{1}{2}\left[1 + \sin\left(\pi\frac{z - \frac{1}{2}(z_{ss}(v) + z_{ba})}{z_{ss}(v) - z_{ba}}\right)\right] \quad . \tag{8.19}$$

Therefore the parameter z_{ba} defines the point where α starts to take non-zero values, and is termed *breakaway displacement.*

It is now time to try to make out some sense from these equations. Suppose that a constant relative velocity v is applied, starting from zero conditions.

1. As far as z remains small ($z < z_{ba}$), then $\alpha = 0$ and the first equation in (8.16) states that $\dot{z} = v$. This describes *presliding elastic* displacement: the (mean) bristle deflection rate equals the relative velocity and the bristle is still anchored to the contact surface.

2. When z exceeds z_{ba}, the mixed *elastic-plastic* regime is entered, where $|\dot{z}| < |v|$.

3. After the transient mixed regime, the first-order equation in (8.16) converges to the equilibrium $\dot{z} = 0$, and steady-state is reached with purely *plastic* bristle displacement. Note that $\dot{z} = 0$ means $z = z_{ss}$. It is now clear why z_{ss} (z at steady-state) has been given this name.

Therefore, the steady-state friction force is $f(v) = \sigma_0 z_{ss}(v)$. In other words, at steady-state the elasto-plastic model converges to the kinetic model (8.15). What interests us is the complex transient that takes place *before* steady-state, which is going to provide our friction sounds with rich and veridical dynamic variations.

Using the elasto-plastic model as the coupling mechanism between two modal resonators, the resulting system is

$$\begin{cases} \ddot{x}_i^{(b)} + g_i^{(b)}\dot{x}_i^{(b)} + \left[\omega_i^{(b)}\right]^2 x_i^{(b)} = \dfrac{1}{m_{il}^{(b)}}(f_e^{(b)} + f) \quad , \quad (i = 1 \dots N^{(b)}) \\ \ddot{x}_j^{(r)} + g_j^{(r)}\dot{x}_j^{(r)} + \left[\omega_j^{(r)}\right]^2 x_j^{(r)} = \dfrac{1}{m_{jm}^{(r)}}(f_e^{(r)} - f) \quad , \quad (j = 1 \dots N^{(r)}) \\ v = \sum_{j=1}^{N^{(r)}} t_{mj}^{(r)}\dot{x}_j^{(r)} - \sum_{i=1}^{N^{(b)}} t_{li}^{(b)}\dot{x}_i^{(b)} \\ \dot{z}(v,z) = v\left[1 - \alpha(z,v)\dfrac{z}{z_{ss}(v)}\right] \\ f = \sigma_0 z + \sigma_1 \dot{z} + \sigma_2 v + \sigma_3 w \quad \text{(friction force)} \end{cases} , \tag{8.20}$$

where for clarity the two objects have been denoted with the superscripts (b) and (r), which stand for “bow” and “resonator”, respectively. The x variables are the modal displacements, while z is the mean bristle displacement. The terms $f_e^{(b)}, f_e^{(r)}$ represent external forces, while the integers $N^{(b)}$ and $N^{(r)}$ are the number of modes for the bow and the resonator, respectively. The relative velocity v has been defined assuming that the interaction occurs at point $l = 1 \dots N^{(b)}$ of the bow and point $m = 1 \dots N^{(r)}$ of the resonator. Note that this system has one degree of freedom (z) more than the impact model given in Eq. (8.14).

8.3.3 Surface texture

Many of the contact sounds we are interested in cannot be convincingly rendered by only using deterministic models. As an example, rolling sounds result from random sequences of micro-impacts between two resonating objects, which are determined by the profile of the contacting surface[4]. Friction modeling also requires the knowledge of some surface texture, in order to synthesize noisy sliding sounds (see Eq. (8.16) and the $\sigma_3 w$ term). In the remainder of this section we therefore address the problem of surface texture rendering by means of fractal processes. Fractal processes are widely used in computer graphics, since they provide surfaces and textures that look natural to a human eye [193]. Since in physics-based modeling there is direct translation of geometric surface properties into force signals and, consequently, into sound, it seems natural to follow the same fractal approach to surface modeling.

Fractals are generally defined [116] as scale-invariant geometric. They are *self-similar* if the rescaling is isotropic or uniform in all directions, *self-affine* if the rescaling is anisotropic or dependent on the direction, as *statistically self-similar* if they are the union of statistically rescaled copies of themselves.

More formally, a one-dimensional fractal process can be defined as a generalization of the definition of standard Brownian motion. As reported in [202], the stochastic process $x = \{x(t), t \geq 0\}$ is *standard Brownian motion* if

[4] Rolling sound design is addressed in detail in chapter 9.

1. the stochastic process x has independent increments;
2. the property
$$x(t) - x(s) \sim \mathcal{N}(0, t - s) \quad \text{for} \quad 0 \leq s < t$$
holds. That is, the increment $x(t) - x(s)$ is normally distributed with mean 0 and variance equal to $(t - s)$;
3. $x(0) = 0$.

The definition of standard Brownian motion can be generalized to the definition of *fractal process*, if the increment $x(t) - x(s)$ is normally distributed with mean 0 and variance proportional to $(t - s)^{2H}$. The parameter H is called *Hurst exponent*, and characterizes the scaling behaviour of fractal processes: if $x = \{x(t), t \geq 0\}$ is a fractal process with Hurst exponent H, then, for any real $a > 0$, it obeys the scaling relation

$$x(t) \overset{\mathcal{P}}{=} a^{-H} x(at) \quad , \tag{8.21}$$

where $\overset{\mathcal{P}}{=}$ denotes equality in a statistical sense. This is the formal definition for *statistical self-similarity*. The $1/f$ family of statistically self-similar processes, also known as $1/f$ *noise*, is defined as having power spectral density $S_x(\omega)$ proportional to $1/\omega^\beta$ for some spectral parameter β related to the Hurst exponent H by $\beta = 2H + 1$. For $\beta = 0$ the definition corresponds to white noise, for $\beta = 2$ Brownian noise is obtained, and for $\beta = 1$ the resulting noise is referred to as pink noise.

The parameter β also relates to the *fractal dimension* [256]. The fractal dimension of a function is related to the roughness of its plot and is exploited in computer graphics to control the perceived roughness [193]. For $1/f$ processes, it is inversely proportional to the Hurst exponent H. Larger values of H correspond to lower values of the fractal dimension and H is proportional to β. Therefore, by increasing β, we will achieve a redistribution of power from high to low frequencies, with an overall smoothing of the waveform.

The problem of generating $1/f$ noise has been treated extensively. One of the most common approaches amounts to properly filtering a white noise source in order to obtain a $1/f$ spectrum. We follow here this approach, and use a model reported in [210] and [55]. The shaping filter is a cascade of N first-order filters, each with a real zero-pole pair. The overall transfer function $H(s)$ in the Laplace domain is the following:

$$H(s) = A \frac{\prod_{i=1}^{N}(s - s_{0i})}{\prod_{i=1}^{N}(s - s_{pi})} \quad , \tag{8.22}$$

where A is a suitable constant.

The fractal noise generator is obtained by properly setting the poles and the zeros of the filters in the cascade [210]. Specifically, the pole and zero frequencies, f_{pi} and f_{0i}, can be computed as

Figure 8.7: Magnitude spectrum of generated fractal noise with $\beta = 1.81$, $h = 2$ (left) and $h = 6$ (right)

functions of the spectral slope β with the following formulas:

$$\begin{aligned} f_{pi} &= -\frac{s_{pi}}{2\pi} = f_{p(i-1)} 10^{\frac{1}{h}} \quad , \\ f_{0i} &= -\frac{s_{0i}}{2\pi} = f_{pi} 10^{\frac{\beta}{2h}} \quad , \end{aligned} \tag{8.23}$$

where f_{p1} is the lowest pole frequency of the filter. Therefore, the lowest limit of the frequency band for the approximation is f_{p1} and the range width is expressed in decades. The density h (density of the poles per frequency decade) can be used to control the error between the target spectrum and the approximated spectrum obtained by white noise filtering. The dependence of the error with respect to filter pole density is discussed in [55]. Figure 8.7 shows a $1/f^{\beta}$ spectrum obtained using the filter (8.22), with two different h values.

The transfer function in the discrete-time domain can be computed with the Impulse Invariant method [177]. It is known that this corresponds to mapping poles and zeros of the transfer function $H(s)$ to poles and zeros of the transfer function $H(z)$ in the discrete-time domain by making the following substitution:

$$s - s_x \longrightarrow 1 - e^{s_x T_s} z^{-1} \quad , \tag{8.24}$$

where T_s is the sampling period and s_x stands for a pole s_{pi} or a zero s_{0i}. The following discrete transfer function is then obtained:

$$H(z) = A' \frac{\prod_{i=1}^{N} 1 - e^{s_{0i} T} z^{-1}}{\prod_{i=1}^{N} 1 - e^{s_{pi} T} z^{-1}} \quad , \tag{8.25}$$

where A' is a normalizing constant. In conclusion, the $1/f^{\beta}$ spectrum is approximated by a cascade of first-order filters, each one with the following discrete transfer function:

$$H^{(i)}(z) = \frac{1 + b_i z^{-1}}{1 + a_i z^{-1}} \quad , \quad \text{with} \quad \left\{ \begin{array}{ll} a_i = e^{-2\pi f_{pi} T}, & b_i = e^{-2\pi f_{0i} T} \\ f_{pi} = f_{p(i-1)} 10^{\frac{1}{h}}, & f_{0i} = f_{pi} 10^{\frac{\beta}{2h}} \end{array} \right. \quad . \tag{8.26}$$

8.4 Implementation and control

As part of the SOb project activities, the low-level sound models described so far have been implemented as real-time modules written in C language for `pd`[5], the open source real-time synthesis environment developed by Miller Puckette and widely used in the computer music community. The modules have been collected into the `interaction_modal` package, downloadable from the Sounding Object web site [6].

When opening the `interaction_modal` folder, one finds a few subdirectories that reflect the object-oriented structure of the plugins:

`resonators`: contains the implementation of resonators described as a bank of modal oscillators, each discretized with the bilinear transformation. External forces can be applied at specified interaction points, each point being described by a set of numbers that weight each mode at that point. Displacement or velocity are returned as outputs from the modal object;

`interactors`: for impact and friction interactions, a function computes the forces to be applied to two interacting resonators using the non-linear equations discussed in the previous section. Details about numerical issues are discussed in appendix 8.A below;

`sound_modules`: contains a subdirectory for each plugin implemented, where the structures and functions required by `pd` are provided. Here, the external appearance (default parameter values, inlets and outlets) of the plugins is also defined.

One critical issue in physically-based sound modeling is parameter estimation and control. Interaction between the user and the audio objects relies mainly upon a small subset of the control parameters. These are the external forces acting on each of the two objects (which have the same direction as the interaction force). In the case of friction, a third high-level parameter is the normal force f_N between the two objects. The remaining parameters belong to a lower level control layer, as they are less likely to be touched by the user and have to be tuned at the sound design level.

Such low-level parameters can be grouped into two subsets, depending on whether they are related to the resonators' internal properties or to the interaction mechanism. Each mode of the two resonating objects is tuned according to its center frequency and decay time. Additionally, the modal gain (inversely proportional to the modal mass) can be set for each resonator mode, and controls the extent to which the mode can be excited during the interaction. The implementation allows position dependent interaction by giving the option to choose any number of interaction points. A different set of modal gains can be set for each point.

A second subset of low-level parameters relates to the interaction force specification, as given in Eqs. (8.11) and (8.16). In certain cases typical parameter values can be found from the literature. Alternatively, they can also be found from analysis of real signals. Parameter estimation techniques are the subject of many studies in automatic control, an extensive discussion of such issues is provided in [7].

[5] http://www.pure-data.org/doc/

[6] http://www.soundobject.org

Figure 8.8: Implementation of the interaction models as `pd` plugins: schematic representation of the sound modules.

This hierarchy for the control parameters is depicted in Figure 8.8, where a schematic representation of the `pd` sound modules is provided. The remainder of this section addresses the phenomenological role of the low-level control parameters.

8.4.1 Controlling the impact model

The impact model has been tested in order to assess its ability to convey perceptually relevant information to a listener. A study on materials [10] has shown that the decay time is the most salient cue for material perception. This is very much in accordance with previous results [137]; however, the physical model used here is advantageous over using a signal-based sound model, in that more realistic attack transients are obtained. The decay times t_{ej} of the resonator modes can therefore be used to "tune" the perceived material of the resonator in a collision with a hammer. See also chapter 4 in this book for more detailed discussion on material perception from recorded and synthesized sounds.

A study on hammer hardness [11] has shown that the contact time t_0 (i.e. the time after which the hammer separates from the resonator) can be controlled using the physical parameters. This is a relevant result, since t_0 has a major role in defining the spectral characteristics of the initial transient. Qualitatively, a short t_0 corresponds to an impulse-like transient with a rich spectrum, and thus provides a bright attack. Similarly, a long t_0 corresponds to a smoother transient with little energy in the high frequency region. Therefore t_0 influences the spectral centroid of the attack

Figure 8.9: Sound spectra obtained when hitting a resonator with a soft mallet (low m_h/k) and with a hard hammer (high m_h/k).

transient, and it is known that this acoustic parameter determines to a large extent the perceived quality of the impact [82]. See also an earlier chapter in this book for a detailed discussion on impact sounds and the perceptual role of the spectral centroid of the attack transient.

An equation has been derived in [11], which relates t_0 to the physical parameters of the model:

$$t_0 = \left(\frac{m^{(h)}}{k}\right)^{\frac{1}{\alpha+1}} \cdot \left(\frac{\mu^2}{\alpha+1}\right)^{\frac{\alpha}{\alpha+1}} \cdot \int_{v_{out}}^{v_{in}} \frac{dv}{(1+\mu v)\left[-\mu(v-v_{in}) + \log\left|\frac{1+\mu v}{1+\mu v_{in}}\right|\right]^{\frac{\alpha}{\alpha+1}}} \quad . \tag{8.27}$$

This equation states that the contact time t_0 depends only on v_{in} and two parameters, i.e. the viscoelastic characteristic μ and the ratio $m^{(h)}/k$. Specifically, the ratio $m^{(h)}/k$ is found to be the most relevant parameter in controlling contact time and consequently the perceived hardness of the impact. Numerical simulations have shown excellent accordance between contact times computed using Eq. (8.27) and those observed in the simulations. Figure 8.9 shows an example of soft and hard impacts, obtained by varying m_h/k.

Due to the physical description of the contact force, realistic effects can be obtained from the model by properly adjusting the physical parameters. Figure 8.10a shows an example output from the model, in which the impact occurs when the resonator is already oscillating: the interaction, and consequently the contact force profile, differs from the case when the resonator is not in motion before collision. This effect can not be simulated using pre-stored contact force profiles (as e.g. in [237]). Figure 8.10b shows an example of "hard collision", obtained by giving a very high value to the stiffness k, while the other model parameters have the same values as in Figure 8.10a. It can be noticed that several micro-collisions take place during a single impact. This is qualitatively in accordance with the remarks about hard collisions by van den Doel et al. [237].

(a) (b)

Figure 8.10: Numerical simulations; (a) impact on an oscillating resonator; (b) micro-impacts in a hard collision. Intersections between the solid and the dashed lines denote start/release of contact.

Symbol	**Physical Description**	**Phenomenological Description**
σ_0	bristle stiffness	affects the evolution of mode lock-in
σ_1	bristle dissipation	affects the sound bandwidth
σ_2	viscous friction	affects the speed of timbre evolution and pitch
σ_3	noise coefficient	affects the perceived surface roughness
μ_d	dynamic friction coeff.	high values reduce the sound bandwidth
μ_s	static friction coeff.	affects the smoothness of sound attack
v_s	Stribeck velocity	affects the smoothness of sound attack
f_N	normal force	high values give rougher and louder sounds

Table 8.1: A phenomenological guide to the friction model parameters.

8.4.2 Controlling the friction model [7]

Similarly to impact, the phenomenological role of the low-level physical parameters of the friction model has been studied. The description given in Table 8.1 can be a helpful starting point for the sound designer.

The triple $(\sigma_0, \sigma_1, \sigma_2)$ (see Eq. (8.16)) define the bristle stiffness, the bristle internal dissipation, and the viscous friction, and therefore affects the characteristics of signal transients as well as the ease in establishing stick-slip motion. The triple (f_c, f_s, v_s) (see Eq. (8.17)) specifies the shape of the steady state Stribeck curve. Specifically, the Coulomb force and the stiction force are related to the normal force through the equations $f_s = \mu_s f_N$ and $f_c = \mu_d f_N$, where μ_s and μ_d are the static

[7] section co-authored with Stefania Serafin

and dynamic friction coefficients[8]. Finally, the breakaway displacement z_{ba} (see equation (8.18)) is also influenced by the normal force. In order for the function $\alpha(v, z)$ to be well defined, the inequality $z_{ba} < z_{ss}(v)\ \forall v$ must hold. Since $\min_v z_{ss}(v) = f_c/\sigma_0$, a suitable mapping between f_N and z_{ba} is

$$z_{ba} = cf_c/\sigma_0 = c\mu_d f_N/\sigma_0 \quad , \quad \text{with} \quad c < 1 \quad . \tag{8.28}$$

By exploiting the above indications on the phenomenological role of the low-level parameters, and their relation to user-controlled parameters, simple interactive applications have been designed which use a standard mouse as the controlling device. Namely, x- and y-coordinates of the pointer are linked to the external force $f_e^{(b)}$ and the normal force f_N, respectively. The applications have been designed using the OpenGL-based `Gem`[9] external graphical library of `pd`.

Braking effects Different kinds of vibrations and sonorities develop within wheel brakes: in the case of rotating wheels slipping sideways across the rails, the friction forces acting at the wheel rim excite transverse vibrations. In the simulation depicted in Figure 8.11a, a wheel driven by the external force $f_e^{(b)}$ rolls on a circular track (a detailed description of rolling sound design is given in the next chapter). When a positive normal force is applied, the wheel is blocked from rolling and the friction model is triggered. Neat stick-slip is established only at sufficiently low velocities, and brake squeals are produced in the final stage of deceleration. The resulting effect convincingly mimics real brake noise.

Wineglass rubbing An excitation mechanism analogous to wheel-brake interaction appears when a wineglass is rubbed around its rim with a moist finger. In this case sound radiates at one of the natural frequencies of the glass and its harmonics. By properly adjusting the modal frequencies and the decay times of the modal object which acts as the resonator, a distinctive glassy character can be obtained. In the example depicted in Figure 8.11b, the rubbing finger is controlled through mouse input. Interestingly, setting the glass into resonance is not a trivial task and requires some practice and careful control, just as in the real world.

Door squeaks Another everyday sound is the squeak produced by the hinges of a swinging door. In this situation, different combinations of transient and continuous sliding produce many squeaks which create a broad range of sonic responses. The example depicted in Figure 8.11c uses two exciter-resonator pairs, one for each of the shutters. In this case the modal frequencies of the objects have been chosen by hand and hear tuning on the basis of recorded sounds. The results are especially convincing in reproducing complex transient and *glissando* effects which are typically found in real door squeaks.

[8] It must be noted that treating f_N as a control parameter is a simplifying assumption, since oscillatory normal force components always accompany the friction force in real systems [5].

[9] http://gem.iem.at/

(a) (b) (c)

Figure 8.11: Interactive applications; (a) a wheel which rolls and slides on a circular track; (b) a moisty finger rubbing a crystal glass; (c) a swinging door, each of the two shutters is linked to a friction module.

8.4.3 Implementation of the fractal noise generator patch

At the time of writing this book chapter the fractal noise generator discussed in section 8.3.3 has not been integrated within the friction model yet. Having done this would (will) allow to control information on surface roughness by controlling the fractal dimension of the noisy component $\sigma_3 w$ in Eq. (8.16).

The fractal noise generator has been implemented independently as a `pd` patch. For convenience in implementation, the shaping filters (8.26) are rewritten as a cascade of *biquads*. Therefore, the cascade is made of $N/2$ second-order filters, each one with the following transfer function (calculated from Eq. (8.26)):

$$\begin{aligned} H^{(i)}(z) = H^{(j)} H^{(j-1)}(z) &= \frac{(1+b_j z^{-1})(1+b_{j-1} z^{-1})}{(1+a_j z^{-1})(1+a_{j-1} z^{-1})} \\ &= \frac{1+(b_j+b_{j-1})z^{-1}+(b_j b_{j-1})z^{-2}}{1+(a_j+a_{j-1})z^{-1}+(a_j a_{j-1})z^{-2}} \quad , \\ &\text{with } j = 2 \cdot i, \; i = 1 \ldots N/2. \end{aligned} \tag{8.29}$$

The `pd` patch of the fractal noise generator has been developed with a modular approach, for future exploitation in physics-based sound design. The most relevant parameter accessible to the user is β, which defines the target $1/f^{\beta}$ spectrum. The number of poles of the filtering cascade can be also set, as well as the frequency of the first pole: these parameters controls the accuracy of the $1/f^{\beta}$ approximation. A snapshot of the patch is given in Figure 8.12.

Figure 8.12: pd patch of the fractal noise generator.

8.A Appendix – Numerical issues

We have shown that the low-level interaction models, impact and friction, are represented through some non-linear coupling between the resonating objects. When the continuous-time systems are discretized and turned into numerical algorithms, the non-linear terms introduce computational problems that require to be solved adequately. This is the topic of this appendix.

8.A.1 Discretization

As already discussed in section 8.2.3, the discrete-time equations for the modal resonator are obtained by using the bilinear transformation. Recalling Eq. (8.6), the "resonator" object is then described in the discrete-time domain by the system

$$\begin{cases} \boldsymbol{x}_j^{(r)}(n) = \boldsymbol{A}_j^{(r)} \boldsymbol{x}_j^{(r)}(n-1) + \boldsymbol{b}_j^{(r)}[y(n) + y(n-1)] \quad , \quad (\text{for} \quad j = 1 \dots N^{(r)}) \\ \boldsymbol{x}^{(r)}(n) = \sum_{j=1}^{N^{(r)}} t_{lj} \boldsymbol{x}_j^{(r)}(n) \quad , \quad (\text{for} \quad j = 1 \dots N^{(r)}) \\ y(n) = f_e^{(r)}(n) - f(n) \end{cases} , \qquad (8.30)$$

where the matrices are given in Eq. (8.7) and $f(n)$ is either the impact force (8.11) or the friction force (8.16). An identical system is written for the "hammer" and the "bow" objects. The state $\boldsymbol{x}^{(r)}(n)$ has been defined assuming that interaction occurs at point j of the resonator.

But how is the interaction force $f(n)$ computed at each time step?

Impact As for the impact force, the missing equations are simply

$$\begin{cases} \begin{bmatrix} x(n) \\ v(n) \end{bmatrix} = \boldsymbol{x}^{(r)}(n) - \boldsymbol{x}^{(h)}(n) \\ f(n) = f(x(n), v(n)) \end{cases} , \qquad (8.31)$$

where f is given in Eq. (8.11). It can be seen that at each time step n the variables $[x(n), v(n)]$ and $f(n)$ have instantaneous mutual dependence. That is, a delay-free non-computable loop has been created in the discrete-time equations and, since a non-linear term is involved in the computation, it is not trivial to solve the loop. This is a known problem in numerical simulations of non-linear dynamic systems. An accurate and efficient solution, called K method, has been recently proposed in [24] and will be adopted here. First, the instantaneous contribution of $f(n)$ in the computation of vector $[x(n), v(n)]$ can be isolated as follows:

$$\begin{bmatrix} x(n) \\ v(n) \end{bmatrix} = \begin{bmatrix} \tilde{x}(n) \\ \tilde{v}(n) \end{bmatrix} + \boldsymbol{K} f(n) \quad \text{with} \quad \boldsymbol{K} = -\left(\sum_{i=1}^{N^{(h)}} t_{mi} \boldsymbol{b}_i^{(h)} + \sum_{j=1}^{N^{(r)}} t_{lj} \boldsymbol{b}_j^{(r)} \right) , \qquad (8.32)$$

where $[\tilde{x}(n), \tilde{v}(n)]$ is a computable vector (i.e. it is a linear combination of past values of $\boldsymbol{x}_j^{(r)}, \boldsymbol{x}_i^{(h)}$ and y). Second, substituting the expression (8.32) in the non-linear contact force equation, and applying the implicit function theorem, $f(n)$ can be found as a function of $[\tilde{x}(n), \tilde{v}(n)]$ only:

$$f(n) = f\left(\begin{bmatrix} \tilde{x}(n) \\ \tilde{v}(n) \end{bmatrix} + \boldsymbol{K} f(n) \right) \quad \overset{\text{K method}}{\longmapsto} \quad f(n) = h(\tilde{x}(n), \tilde{v}(n)) \quad . \qquad (8.33)$$

Summarizing, if the map $f(n) = h(\tilde{x}(n), \tilde{v}(n))$ is known, then the delay-free loop in the computation can be removed by rewriting the algorithm as

```
for   $n = 1\ldots$ samplelength
      Assign    $f(n) = 0$
      Compute   $\boldsymbol{x}_i^{(h)}(n)$ $(i = 1\ldots N^{(h)})$, and $\boldsymbol{x}_j^{(r)}(n)$ $(j = 1\ldots N^{(r)})$
      Compute   $\tilde{x}(n)$, $\tilde{v}(n)$, and $f(n) = h(\tilde{x}(n), \tilde{v}(n))$
      Update    $\boldsymbol{x}_i^{(h)}(n) = \boldsymbol{x}_i^{(h)}(n) + \boldsymbol{b}_i^{(h)} f(n)$ $(i = 1\ldots N^{(h)})$
      Update    $\boldsymbol{x}_j^{(r)}(n) = \boldsymbol{x}_j^{(r)}(n) - \boldsymbol{b}_j^{(r)} f(n)$ $(j = 1\ldots N^{(r)})$
end
```

Friction The numerical implementation for frictional contact (8.16) is slightly more complicated, because of the additional degree of freedom z. The dynamic equation for $\dot{z}$ is again discretized using the bilinear transformation. Since this is a first order equation, discretization by the trapezoid rule is straightforward:

$$\begin{aligned} z(n) = &\quad z(n-1) + \int_{(n-1)T_s}^{nT_s} \dot{z}(\tau) d\tau = \\ \Rightarrow z(n) \approx &\quad z(n-1) + \frac{T_s}{2}\dot{z}(n-1) + \frac{T_s}{2}\dot{z}(n) \quad . \end{aligned} \tag{8.34}$$

Therefore, the missing equations in the coupled numerical system are

$$\left\{ \begin{array}{l} \left[\begin{array}{c} x(n) \\ v(n) \end{array} \right] = \boldsymbol{x}^{(r)}(n) - \boldsymbol{x}^{(b)}(n) \\ z(n) = z(n-1) + \frac{T_s}{2}\dot{z}(n-1) + \frac{T_s}{2}\dot{z}(n) \\ \dot{z}(n) = \dot{z}(v(n), z(n)) \\ f(n) = f(z(n), \dot{z}(n), v(n), w(n)) \end{array} \right. \quad , \tag{8.35}$$

where $\dot{z}(v, z)$ and $f(z, \dot{z}, v, w)$ are given in Eq. (8.16). Again, it can be seen that at each time step n the variables $[v(n), z(n)]$ and $\dot{z}(n)$ have instantaneous mutual dependence. Again, the K method [24] will be adopted in order to solve this problem. To this end, the instantaneous contribution of $\dot{z}(n)$ in the computation of vector $[x(n), v(n)]$ must be isolated so that the K method can be applied on the non-linear function $\dot{z}(v, z)$:

$$\left[\begin{array}{c} v(n) \\ z(n) \end{array} \right] = \left[\begin{array}{c} \tilde{v}(n) \\ \tilde{z}(n) \end{array} \right] + \boldsymbol{K}\dot{z}(n) \quad , \tag{8.36}$$

then

$$\dot{z}(n) = \dot{z}\left(\left[\begin{array}{c} \tilde{v}(n) \\ \tilde{z}(n) \end{array} \right] + \boldsymbol{K}\dot{z}(n) \right) \quad \overset{\text{K method}}{\longmapsto} \quad \dot{z}(n) = h(\tilde{v}(n), \tilde{z}(n)) \quad . \tag{8.37}$$

where $\tilde{v}$ and $\tilde{z}$ are –as above– computable quantities. From Eq. (8.34), the element $\boldsymbol{K}(2)$ is easily found as $\boldsymbol{K}(2) = T_s/2$, while $\tilde{z}(n) = z(n-1) + T_s/2 \cdot \dot{z}(n-1)$. Finding $\boldsymbol{K}(1)$ is less straightforward, since the friction force itself depends explicitly upon v. Recalling that

$$v(n) = \sum_{j=1}^{N^{(r)}} t_{lj} \dot{x}_j^{(r)}(n) - \sum_{j=1}^{N^{(b)}} t_{mi} \dot{x}_i^{(b)}(n) \quad , \tag{8.38}$$

and substituting here the discrete-time equations (8.30) for $\dot{x}_j^{(r)}(n)$ and $\dot{x}_i^{(b)}(n)$, a little algebra leads to the result

$$\begin{aligned} v(n) = {} & \frac{1}{1+\sigma_2 b} \left\{ \sum_{j=1}^{N^{(r)}} t_{lj} \left\{ \dot{\tilde{x}}_j^{(r)}(n) + \boldsymbol{b}_j^{(r)}(2)[f_e^{(r)}(n) - \sigma_0 \tilde{z}(n)] \right\} - \right. \\ & \left. - \sum_{i=1}^{N^b} t_{mi} \left\{ \dot{\tilde{x}}_i^{(b)}(n) + \boldsymbol{b}_i^{(b)}(2)[f_e^{(b)}(n) + \sigma_0 \tilde{z}(n)] \right\} \right\} - \\ & - \frac{b}{1+\sigma_2 b} \left(\sigma_0 \frac{T_s}{2} + \sigma_1 \right) \dot{z}(n) = \\ = {} & \tilde{v}(n) + \boldsymbol{K}(1) \dot{z}(n) \quad , \end{aligned} \tag{8.39}$$

where the quantities $\dot{\tilde{x}}^{(r,b)}$ are the “computable” part of the modal velocities (i.e., they are computed from modal resonator equations (8.30) with $f(n) = 0$), and the term b is defined as $b = \left[\sum_{i=1}^{N^{(b)}} t_{mi} \boldsymbol{b}_i^{(b)}(2) + \sum_{j=1}^{N^{(r)}} t_{lj} \boldsymbol{b}_j^{(r)}(2) \right]$. The quantities $\tilde{v}$ and $\boldsymbol{K}(1)$ are defined in Eq. (8.39) in an obvious way. Having determined the $\boldsymbol{K}$ matrix, the K method can be applied and the algorithm can be rewritten as

```
for    $n = 1 \ldots$ samplelength
       Assign     $f(n) = 0$
       Compute    $\boldsymbol{x}_i^{(b)}(n)$  $(i = 1 \ldots N^{(b)})$, and $\boldsymbol{x}_j^{(r)}(n)$  $(j = 1 \ldots N^{(r)})$
       Compute    $\tilde{v}(n)$, $\tilde{z}(n)$, and $\dot{z}(n) = h(\tilde{v}(n), \tilde{z}(n))$
       Compute    $v(n) = \tilde{v}(n) + \boldsymbol{K}(1)\dot{z}(n)$, $z(n) = \tilde{z}(n) + \boldsymbol{K}(2)\dot{z}(n)$, and $f(n)$
       Update     $\boldsymbol{x}_i^{(b)}(n) = \boldsymbol{x}_i^{(b)}(n) + \boldsymbol{b}_i^{(b)} f(n)$  $(i = 1 \ldots N^{(b)})$
       Update     $\boldsymbol{x}_j^{(r)}(n) = \boldsymbol{x}_j^{(r)}(n) - \boldsymbol{b}_j^{(r)} f(n)$  $(j = 1 \ldots N^{(r)})$
end
```

8.A.2 The Newton-Raphson algorithm

Two choices are available for efficient numerical implementation of the K method. The first choice amounts to pre-computing the new non-linear function h off-line and storing it in a look-up table. One drawback is that when the control parameters (and thus the $\boldsymbol{K}$ matrix) are varied over time, the function h needs to be re-computed at each update of $\boldsymbol{K}$. In such cases, an alternative and more convenient approach amounts to finding h iteratively at each time step, using the Newton-Raphson method. This latter approach is adopted here. Since most of the computational load in the numerical system comes from the non-linear function evaluation, the speed of convergence (i.e. the number of iterations) of the Newton-Raphson algorithm has a major role in determining the efficiency of the simulations.

Using the Newton-Raphson method for computing h means that at each time step n the value $h(n)$ is found by searching a local zero of the function

$$g(h) = \begin{cases} f\left(\begin{bmatrix} \tilde{x} \\ \tilde{v} \end{bmatrix} + \boldsymbol{K}h\right) - h \quad \text{(impact)} \\ \dot{z}\left(\begin{bmatrix} \tilde{v} \\ \tilde{z} \end{bmatrix} + \boldsymbol{K}h\right) - h \quad \text{(friction)} \end{cases} . \tag{8.40}$$

The Newton-Raphson algorithm operates the search in this way:

```
$h_0 = h(n-1)$
$k = 1$
while (err < Errmax)
        Compute    $g(h_k)$ from Eq. (8.40)
        Compute    $h_{k+1}$ as $h_{k+1} = h_k - \frac{g(h_k)}{g'(h_k)}$
        Compute    err = abs($h_{k+1} - h_k$)
        $k = k + 1$
end
$h(n) = h_k$
```

Therefore, not only the function $g(h)$ but also its derivative $g'(h)$ has to be evaluated at each iteration. As for the impact, this is found as a composite derivative:

$$\frac{dg}{dh} = \frac{\partial f}{\partial x}\boldsymbol{K}(1) + \frac{\partial f}{\partial v}\boldsymbol{K}(2) - 1 \quad , \tag{8.41}$$

where (recalling Eq. (8.11))

$$\begin{aligned} \frac{\partial f}{\partial x} &= \alpha x^{\alpha-1}\left[k + \lambda v\right] \quad , \\ \frac{\partial f}{\partial v} &= \lambda x^{\alpha} \quad . \end{aligned} \tag{8.42}$$

As for friction, the computation of $g'(h)$ is slightly lengthier. Again, it is done in successive steps as a composite derivative. First step:

$$\frac{dg}{dh} = \frac{\partial \dot{z}}{\partial v}\boldsymbol{K}(1) + \frac{\partial \dot{z}}{\partial z}\boldsymbol{K}(2) - 1 \quad . \tag{8.43}$$

Second step (recalling Eq. (8.16)):

$$\begin{aligned}\frac{\partial \dot{z}}{\partial v} &= 1 - z\left[\frac{(\alpha + v \cdot \partial\alpha/\partial v) z_{ss} - \alpha \cdot v \cdot dz_{ss}/dv}{z_{ss}^2}\right] \quad ,\\ \frac{\partial \dot{z}}{\partial z} &= -\frac{v}{z_{ss}}\left[z\frac{\partial \alpha}{\partial z} + \alpha\right] \quad .\end{aligned} \tag{8.44}$$

Third step (recalling Eqs. (8.18, 8.19)):

$$\frac{\partial \alpha}{\partial v} = \begin{cases} \frac{\pi}{2}\cos\left(\pi\frac{z - \frac{z_{ss}+z_{ba}}{2}}{z_{ss} - z_{ba}}\right)\frac{\frac{dz_{ss}}{dv}(z_{ba} - z)}{(z_{ss} - z_{ba})^2} \quad , & \begin{array}{l}(z_{ba} < |z| < z_{ss})\ \&\\ (\mathrm{sgn}(v) = \mathrm{sgn}(z))\end{array} \\ 0 \quad , & \text{elsewhere}\end{cases} \quad , \tag{8.45}$$

$$\frac{\partial \alpha}{\partial z} = \begin{cases} \frac{\pi}{2}\cos\left(\pi\frac{z - \frac{z_{ss}+z_{ba}}{2}}{z_{ss} - z_{ba}}\right)\frac{1}{z_{ss} - z_{ba}} \quad , & \begin{array}{l}(z_{ba} < |z| < z_{ss})\ \&\\ (\mathrm{sgn}(v) = \mathrm{sgn}(z))\end{array} \\ 0 \quad , & \text{elsewhere}\end{cases} \quad . \tag{8.46}$$

Last step (recalling Eq. (8.17)):

$$\frac{dz_{ss}}{dv} = -\mathrm{sgn}(v)\frac{2v}{\sigma_0 v_s^2}(f_s - f_c)e^{-(v/v_s)^2} \quad . \tag{8.47}$$

Computing these terms from the last step to the first step, the derivative $g'(h)$ can be obtained.

In order to develop a real-time model, it is essential that the number of iterations for the Newton-Raphson algorithm remains small in a large region of the parameter space. To this end, analysis on the simulations has to be performed, where model parameters are varied over a large range. Such analysis shows that in every conditions the algorithms exhibit a high speed of convergence. More precisely, in the case of the impact the number of iterations is observed to be never higher than four, even when the Newton-Raphson algorithm is given extremely low tolerance errors (Errmax$\sim 10^{-13}$). As for friction, the number of iterations remains smaller than seven.

Chapter 9

High-level models: bouncing, breaking, rolling, crumpling, pouring

Matthias Rath and Federico Fontana
Università di Verona – Department of Computer Science
Verona, Italy
rath@sci.univr.it, fontana@sci.univr.it

9.1 Sound design around a real-time impact model — common basics

Collisions between solid objects form a wide class of sonic events in everyday listening. Impacts are basic sub-events of acoustic significance for a large variety of scenarios as e.g. bouncing, rolling, sliding, breaking or machine actions. The extraction/estimation of *structural invariants* [92], i.e. attributes of involved objects such as size, shape, mass, elasticity, surface properties or material, as well as *transformational invariants*, such as velocities, forces and position of interaction points, from impact-based sounds is common experience. Several psychoacoustic studies exist on the topic, focusing on different aspects, starting from different assumptions and following various strategies, with consequently different results. Lutfi and Oh [163] examine material perception from the resonance behavior, more exactly impulse responses, of "ideal" bars of fixed shape, ignoring internal friction. Klatzky, Pai and Krotkov [137] on the other hand test impulse responses for a possible shape independent acoustic material constant, based exactly on internal friction, not surprisingly gaining somewhat opposite results. Van den Doel and Pai [238] use the latter experiences to lay out a broader method of rendering sounds of hit objects, under additional consideration of interaction position. Common to all these studies is the focus on structural invariants internal to the resonating objects: the significance of a complex transient behavior, reflecting surface depending details of an impact or force and velocity, is not taken into account. In fact, little is known about the perception of transients. Freed [82] developed acoustic parameters related to perceived hardness for a set of recorded percussive sounds; we are though not aware of a respective strategy for the synthesis of impact sounds. Here lies a main point of our basic underlying impact algorithm. Besides the flexible *modal description* of the colliding objects, that forms the basis, and in turn allows immediate use, of existing studies, we consider a physical model of an impact event. This allows us to synthesize convincing impact transients depending on physically meaningful parameters, despite the (current) lack of satisfactory underlying models of their auditory perception. Intuitive handling and the real-time implementation of different objects allow practical access, based on, but not restricted by the boundaries of theoretical foundations.

Finally, higher-level control structures are used to both investigate and exploit the perceptual significance of (statistical or regular) temporal distribution of impacts and their varying attributes. Warren and Verbrugge's investigation [251] of auditory perception of breaking and bouncing events under special consideration of temporal patterns, is the main theoretical psychophysical orientation point for our modeling efforts.

9.1.1 The fundamental physics-based model

In contrast to several studies of contact sounds of solid bodies that focus on the *resonance* behavior of interacting objects and widely ignore the transient state of the event, our algorithm is based on a physical description of impact interaction processes (see chapter 8). This physical model involves a degree of simplification and abstraction that implies efficient implementation as well adaption to a broad range of impact events.

We consider two resonating objects and assume that their interaction depends on the difference x of two (one-dimensional) variables connected to each object. In the standard case of examined movements in one spatial direction, x is the distance variable in that direction. Possible simultaneous interaction along other dimensions are "sorted out" at this stage. This leads to a compact efficient algorithm that strikes the main interaction properties. The impact force f is explicited as a nonlinear term in x and $\dot{x}$, according to Eq. (8.11). There, k is the elasticity constant, i.e. the hardness of the impact. α, the exponent of the non-linear terms shapes the dynamic behavior of the interaction (i.e. the influence of initial velocity), while λ weighs the dissipation of energy during contact, accounting for friction loss. The instantaneous cross relationship of resonator states is solved through a numerical strategy [24], that carefully avoids distorting artifacts; a simpler linear force term can be chosen alternatively, that trades richness in detail for reduced computational cost. Interaction equations are formulated in a way that enables modular connection of numerous different resonators (and also different interactors — besides the linearized version of the impact interactor a friction model has been implemented using the same existing structure).

All interacting, resonating objects used in the following design examples are built under the premises of modal synthesis [2]. This formulation supports particularly well our main design approach for its physical generality and, at the same time, for its intuitive acoustic meaning. In most examples, one resonator (referred to in the following as "striker" or "hammer") for practical and computational convenience is furthermore simplified to an inertial point mass, which is the special

(a) (b)

(c)

Figure 9.1: Inertial mass hitting a resonator with three resonant modes/frequencies . The objects are in contact when the parabola-like trajectory is above the oscillating trajectory. The distance variable is depicted below (shifted to the boundary of the window). The elasticity constant, i.e. the hardness of the contact(surface) is increased from (a) to (c); note that for harder collisions several "micro-impacts" occur, and that high frequency components during contact increase (best to note from the difference curve).

modal resonator with only one resonant mode of frequency 0 and infinite decay time (undamped). Alternative resonating objects, e.g. realized using digital waveguides, can be easily integrated.

One strong point of the described impact model lies in its capability to produce convincing and, if intended, realistic transients that are capable of expressing various impact attributes such as hardness/elasticity, "stickiness", mass proportions and velocity. Currently, signal based methods of synthesis can not satisfy these needs, due to lack of knowledge about perception and production of such transients. On the other hand, as opposed to wavetable techniques, the algorithm can be instantiated with an infinite variety of interaction- and resonator properties, which can be easily tuned to attributes of ecological hearing (material, shape, impact position or surface properties a.s.); also the exact form of each impact depends on the actual state of the involved resonators (see Figure 9.2).

(a) (b)

Figure 9.2: Impacts with identical parameters as in Figure 9.1 (c) on a larger scale; the "hammer" trajectory is shifted as the distance trajectory for a clearer view. Note that the difference between (a) and (b) is only due to the different states of object2 at first contact. All other parameters, including the initial hammer velocity are equal.

Figures 9.1 and 9.2 show some impacts realized with the algorithm, in its application to an inertial mass ("hammer") and a three-mode resonator. The modes of the latter (in frequency, decay and level) were originally set according to the theoretic description of a vibrating bar, but adapted for a possibly clear graphical display. Figure 9.1 focuses on the transient behavior during contact, at different elasticity constants k. Figure 9.2 gives an overview including several free vibrating cycles of the second resonator. It demonstrates how different impacts, even under identical structural invariants result in different transient states and different weighting of modes. This effect is of strong acoustic importance for dense temporal grouping of impacts (e.g. "ringing", or fast pseudo-periodic machine-like sounds).

The influence of mass relations on the impact shape is demonstrated in figures 9.3 and 9.4. While for a small hammer mass an impulse response can be a good approximation, there is no conventional technique to describe/generate impacts for increasing relative hammer mass.

Figure 9.5 depicts interactions with strong non-linear terms ($\alpha = 2.8,\ \lambda \simeq 20$). Again the shape of the signal during contact is not easily analyzed with conventional methods, but is of strong acoustic character.

9.1.2 Resonator attributes

The adjustment of modal parameters is intuitive and can be based on, but is not restricted to, theoretical results or specific practical strategies: [9] contains an overview of the background and practical handling of resonant modes. Some of the described techniques were used to apply the impact algorithm with the resonant behavior of 2D and 3D resonating shapes [208], and in the implementation of a virtual drum [168]. Avanzini and Rocchesso tested sounds realized with a preliminary version of the algorithm for their possible assignment to material attributes [10]. The experimental activities in sound perception have shown that shape variations have subtle con-

(a) (b)

(c)

Figure 9.3: Hard impacts with decreasing hammer mass (from (a) to (c)); the initial velocity is adapted to compensate for similar amplitudes / energy passing. For low hammer masses the impact curve resembles a short impulse; such contacts can thus be more successfully replaced with filtered impulses (e.g. short noise bursts), as done before [238]. In the following free decay of the modal resonator higher frequency components accordingly gain in weight.

sequences on the perceived sonic imagery. On the other hand, acoustic characteristics of materials are undoubtedly a promising basis for expressive sound design.

In fact, it seems possible to reliably express a material attribute through appropriate model parameter configurations. Important here is the capability to reflect material attributes in both resonance (as done in previous approaches) and interaction properties of the objects in contact. Further higher-level modeling efforts strongly rest and rely on a convincing basis of material attributes (in particular "breaking" scenarios: see section 9.3). From a current pragmatic standpoint it has to be noted that existing studies of material perception start from somewhat opposite assumptions and consequently lead to different results. A linear dependency of decay time over frequency, expressed through a material-specific coefficient of internal friction, as developed in the the well-known position paper by Wildes and Richard [253], seemed to be the most effective approach to acoustic expression of material. Figure 9.6 shows a part of a `pd` patch that specifies

(a) (b)

Figure 9.4: Very high relative hammer masses. The elasticity constant is slightly lower (softer) than for Figure 9.3, all other parameters (except adapted initial hammer velocity) are equal. For many physical resonators these examples should exceed the usability of the linear description. For sound (b) a short low frequency impulse during impact can be perceived, while in the following decay the low frequency mode is hardly present.

decay time parameters for modal frequencies after an internal friction parameter and an additional coefficient representing external frictional losses. This method, that has been used [237] and supported through psychoacoustic testing [137] before, is here completed with the described physically-based model of the impact itself. The resulting capability to include further (material-/surface-specific) interaction parameters, such as hardness of contact or "stickiness" fundamentally contributes toward expressivity and realism. Of course these examinations would open up a wide field for systematic testing. One should keep in mind that the linear decay/frequency dependence is only one possible approximation, and psychoacoustic studies e.g. also show a slight influence of frequency ranges on material impression [137].

On the background of a missing overall closed theory of the auditory capabilities and mechanisms of material perception the intuitive accessibility of model parameters provides a chance for sound design: keeping in mind diverging starting points and results of existing studies, the exploitation of different approaches, as well as orientation through immediate subjective feedback, can be a rewarding challenge when reaching for different design goals.

9.2 Bouncing

9.2.1 Preparatory observations

Short acoustic events like impacts can strongly gain or change in expressive content, when set in an appropriate temporal context. One example is the grouping of impacts in a "bouncing" pattern. The physical model underlying our impact algorithms allows the input of an external force term. A bouncing process can be simply achieved with an additional constant term representing

(a) (b)

(c)

Figure 9.5: Contribution of nonlinear terms. Here the contact surface is “sticky”, resulting in a complex and acoustically characteristic transient. (b) and (c) use heavier hammer and slightly harder contact surface. Again, differences in (b) and (c) result from different initial states only; contact parameters are unchanged.

gravity. Figure 9.7 shows a resulting trajectory. It can be surprising how this acoustic grouping of single events, which in isolation do not bear a strong ecological meaning, creates an immediate characteristic association: a bouncing ball.

The above way of generating a temporal pattern is not satisfactory in our context. Due to the physical description, the exact (accelerating) tempo of bouncing is coupled to the impact parameters. Simplifications on the elementary sound level necessarily affect the higher level pattern, demanding compensation. From a standpoint of cartoonification the low-level physical model is “too realistic”. In addition to this unhandiness, the one-dimensionality of the model leads to a regular pattern as it occurs in (3D) reality only for perfect spherical objects or special, highly restricted, symmetric situations. These restrictions led to the development of a “bouncer” control structure, that explicitly creates typical patterns of falling objects. Underlying considerations are sketched in the following.

Figure 9.6: Example of the impact module with a resonator tuned according to the theory of a thin bar. The subpatch calculates the decay times, according to the partial frequencies and coefficients of external and (material-specific) internal friction.

9.2.2 A macroscopic view on bouncing objects

The kinetic energy of a falling solid object can be written as the sum of three terms depending on the vertical and horizontal velocity of its center of mass and its rotation around an axis passing through the center of mass. Of course here, kinetic energy of inner vibration is assumed negligibly small in comparison to these macroscopic components. In a vertical gravity field and under further neglection of friction in surrounding gas, the latter two "horizontal" and "rotational" terms stay constant while the object is not in contact with the ground (or other solids). Only energy related to the vertical movement is translated to or from (for up- or downward movements) potential energy in the gravity field due to the vertical acceleration, that affects only the respective vertical velocity of the center of mass.

We start with the analysis of the free movement of a bouncing object that is reflected at the ground at time $t = 0$ with an upward vertical velocity $v(0) = v_0$ of its center of mass. For a

(a) (b)

Figure 9.7: An inertial mass "bouncing" on a two-mode resonator. (b) focuses on the final state of the process: the two interacting objects finally stay in constant contact, a clear difference to simply retriggering.

constant gravity acceleration g, v decreases according to

$$v(t) = v_0 - g \cdot t, \;\; g > 0 \quad , \tag{9.1}$$

as the center of mass performs a movement "parabolic in time" with its momentary height x described by

$$x(t) = v_0 \cdot t - \frac{g}{2} \cdot t^2 \quad . \tag{9.2}$$

During this free re-bounce between two reflections at the ground the vertical kinetic energy term $E_{kin}(t) = \frac{M}{2} \cdot v^2(t)$, M denoting the overall mass, first decays to 0 along with $v(t)$ until height x and potential energy reach a maximum. While the object falls down again, its potential energy is re-transferred to E_{kin}. Both terms reach its initial values together with x, concurrently the velocity returns to its original absolute value but in opposite (downward) direction $v(t_{return}) = -v_0$. For the bouncing interval follows

$$t_{return} = \frac{2}{g} \cdot v_0 \quad , \tag{9.3}$$

i.e. proportionality to the vertical velocity after reflection (as a reproof one can check that $x(t_{return}) = 0$ using the expression given above).

Next, the loss of macro-kinetic energy in friction and microscopic (a.o. acoustic) vibration with each reflection is looked at as the basic (and, since we neglect friction forces in surrounding gas, ruling) principle behind the process of a decaying bouncing movement. First, horizontal and rotational movements are neglected, assumed independent of the vertical movement, as can be approximately true for the case of a perfectly symmetric (e.g. spherical) bouncing object. Energy

Figure 9.8: A non-spherical object reflected at ground in two different states. Here, a particularly clear example is chosen, a “stick” with its mass concentrated at both ends. The rotation is in both case about an axis parallel to the ground.

and velocities here coincide with their respective vertical components. The amount of energy-“loss” during reflection is in exactness generally different for each impact. This can be seen e.g. from Figure 9.2, where different interaction patterns are displayed, between two identical objects in identical macroscopic but varying microscopic preconditions. Only such elementary simulations can quantify energy transfer at this level of detail. An approximate assumption though is a loss of energy with each bounce proportional to the remaining kinetic energy; this applies e.g. to the ideal case of a damped linear collision force and a fixed, i.e. infinitely inert and stiff “reflector”, which is a good (macroscopic, of course *not* acoustic) approximation for many typical situations of bouncing. Rewriting, we receive a relation of kinetic energy terms E_{pre} and E_{post}, before and after each reflection,

$$E_{post} = C \cdot E_{pre}, \; C < 1 \quad , \tag{9.4}$$

where C is constant for the specific bouncing-scenario. Kinetic energy and velocity at each reflection, as well as the temporal bouncing intervals t_{int} then follow exponentially decaying, in the number of reflections n, terms

$$E(n) = C^n \cdot E_0, \;\; v(n) = \sqrt{C}^n \cdot v_0, \;\; t_{int}(n) = \sqrt{C}^n \cdot t_{int}(0) \quad . \tag{9.5}$$

The implementation of this basic scheme in fact delivered very convincing results in comparison to the afore-described implicit pattern simulation. In Figure 9.7 one can see the strong principal similarity of a bouncing-trajectory as gained from the detailed (one-dimensional) physics-based simulation with the exponential decay behavior derived above. Of course, the final state of the interaction is not preserved with the realism of the implicit, strictly physical-model-based simulation; however, in scenarios labeled “bouncing” the segment in question is of very small amplitude in relation to the initial impacts, so that this difference is hardly noticeable here.

So far, the possible transfer of energy between vertical, horizontal and rotational components with each reflection has been neglected, leading to the pattern that is typical for perfectly round bouncing objects. For irregularly shaped objects this assumption is not applicable, as e.g. everyday experience tells (see also Figure 9.8). This is the reason for the occurrence of individual, often irregular patterns. Again, in general the exact movement in the non-spheric case can only be simulated through a detailed solution of the underlying differential equations. This strategy is highly demanding in terms of complexity of implementation and computational cost and would not make sense in our context of real-time interactivity and cartoonification: it is questionable, how precisely shapes of bouncing objects (except for sphericity) can be recognized acoustically? However, some rough global analysis of bouncing movements lays a basis for the expression of shape properties through an extension of the explicit pattern generation process developed so far. Of the three velocity and respective energy terms after one reflection, only the vertical one (connected to the maximum height of the following bounce) contributes a simple term to the following impact interval and velocity. The horizontal movement has no influence on either, if we admit friction forces only in the direction of the impact, perpendicular to the surface, as in our model of interaction. This neglection of parallel (to the surface) friction is in good acoustic accordance with a wide range of real contact sounds. Finally, the rotation of the bouncing object can in- or decrease (or neither of both) the velocity of the following impact, depending on the momentary angle and direction of rotation. Rotation can also shorten or lengthen the following bouncing interval, since for non-spherical objects the effective height of the center of mass can vary with each reflection, depending on the state of rotation (the angle). The latter effect is seen to be rather subtle, except for situations where the freedom of rotation is limited through small heights of bounces – stages of the scenario that usually call for separate modeling stages, as discussed below. Generally, we can say that rotational and horizontal energy terms, which add up with the vertical term to an approximately exponentially decaying overall energy, lead to — irregularly, quasi randomly — shorter temporal intervals between bounces, bounded by the exponential decay behavior explained above. Rotational movement is also responsible for deviations of the effective impact velocities from the exponential pattern — again basically within the maximal boundaries of the spherical case. Also, the effective mass relations for each impact, but more importantly impact position, vary due to rotation. Consideration of these deviations, especially of the latter effect through respective modulation of modal weights, shows to be of strong perceptual significance.

Very important can be static stages in bouncing-movements, as they occur also for non-spherical, even unsymmetrical, objects, when the rotational freedom is strongly bounded during the final decay of the bouncing-height. In these cases, familiar e.g. from disks or cubes, the transfer of energy between the vertical, horizontal and rotational terms can take place in regular patterns, closely related to those of spherical objects. This phenomenon is exploited in some modeling examples; often however, such movements include rolling aspects, suggesting a potential of improvement through integration of rolling models. A very prominent sound example with an initial "random" and a final regular stage is that of a falling coin.

9.2.3 A cartoon model of bouncing sounds

Summing up the observations of the previous subsection, the "bouncer" patch generates temporal patterns of impact velocities triggered by a starting message. Control parameters are:

1. the time between the first two reflections, representing the initial falling height or, equivalently, falling velocity;

2. the initial impact velocity;

3. the acceleration factor. It is the quotient of two following maximal "bounce-intervals" and describes the amount of microscopic energy loss/transfer with each reflection and thus the speed of the exponential time sequence;

4. the velocity factor, defined analogously;

5. two parameters specify the range of random deviation below the (exponentially decaying) maxima for temporal intervals and impact velocities, respectively. The irregularity/sphericity of an object's shape is modeled in this way;

6. a threshold parameter controls, when the accelerating pattern is stopped, and a "terminating bang" is sent, that can e.g. trigger a following stage of the bouncing process.

Note that the first three parameters should be equal for a spherical object under the above assumed simplifications (see Eq. (9.5)), while in exactness being varied (in dependence of actual impact velocities) in the general case. In a context of cartoon-based auditory display they can be effectively used in a rather intuitive free fashion.

9.3 Breaking

The auditory perception of breaking and bouncing events is examined in a study by Warren and Verbrugge [251]. It is shown, that sound artifacts, created through layering of recorded collision sounds, were identified as bouncing or breaking scenarios, depending on their homogeneity and the regularity and density of their temporal distribution. Also, a short initial noise impulse is shown to contribute to a "breaking" impression.

These results can be effectively exploited and expanded by higher-level sound models, making use of the "impact" module. A first trial is based on Warren and Verbrugge's consideration, that a breaking scenario contains the sub-events of emitted, falling and re-bouncing fragments. Some further thoughts strongly help on successful modeling: typical fragments of rupture are of highly irregular form and are rather inelastic. Consequently, breaking can not be deduced from bouncing movements. In fact, fragments of, e.g., broken glass rather tend to "nod", i.e. perform a decelerating instead of accelerating movement. (The integration of "rolling" and "sliding" (friction) modules is a next planned promising step, on these presumptions.) It is secondly important to keep in mind that emitted fragments mutually collide, and that the number of such mutual collisions rapidly decreases, starting with a massive initial density; those collisions do not describe bouncing patterns at all. Following these examinations a "breaking" model was realized by use of the bouncer with high values of "randomness", and a quickly decreasing temporal density, i.e. a time-factor set "opposite" to the original range for bouncing movements. Again, the increase in expressivity through careful higher-level control, here realized through a small extension of the bouncer, the "dropper", which admits augmenting time-factors, i.e. > 1, can be surprising. Even sounds realized with only one impact-resonator pair can produce a clear breaking-impression. Supporting Warren and Verbrugge's examination, a short noise impulse added to the attack portion of the pattern fortified the breaking character.

As another insight during the modeling process, several sound attributes showed to be important. Impacts grouped in a temporal pattern as explained above, seem to be less identifiable as a breaking event, when tuned to a metallic character in their modal settings; this may correspond to the fact that breaking metal objects are rather far from everyday experience. Also, extreme mass relations of "striker" and struck resonator in the impact settings led to more convincing results. Again, this is in correspondence with typical situations of breakage: a concrete floor has a practically infinite inertia in comparison to a bottle of glass. These mass relations are reflected in distinct attack transients (see section 9.1.1, e.g. Figure 9.3), and the phenomenon is another hint on the advantage of the physics-based low-level impact algorithm. Certainly, these informal experiences could be subject of systematic psychophysical testing.

9.4 Rolling

Among the various common mechanical interactions between solid objects, "rolling" scenarios form a category that seems to be characteristic also from the auditory viewpoint: everyday experience tells that the sound produced by a rolling object is often recognizable as such, and in general clearly distinct from sounds of slipping, sliding or scratching interactions, even of the same objects. This may be due to the nature of rolling as a continuous interaction process, where the mutual force on the involved objects is described as an impact without additional perpendicular friction forces.

Besides being characteristic, rolling-sounds carry strong ecological information: in addition to the (inner) resonance characteristics of the involved objects (which depend on shape, size and material), further detailed attributes of their form or surface are as well acoustically reflected as *transformational* [92] attributes, such as velocity, gravity or acceleration/deceleration. This suggest acoustic modeling of rolling to be a rewarding goal under the various demands of auditory display.

The development of an expressive real-time sound model of rolling from physical, acoustical and implementational presumptions is described in the following.

9.4.1 Rolling interaction with the impact-model as lowest-level building block

In contrast to slipping, sliding or scratching actions, the interaction force on the two objects involved in a simple rolling-scenario (the rolling object and the "plain" to roll on) is basically

Figure 9.9: Sketch of the fictional movement of a ball, perfectly following a surface profile s(x). Relative dimensions are highly exaggerated for a clearer view. Note that this is **not** the de-facto movement; this idealization is used to derive the offset-curve to be used by the impact-model.

perpendicular to the contact surface (the macroscopic mean curve), pointing along the connection line of the momentary point of contact and the "center of gravity of the rolling object"[1]. This raises the promise that our physics-based impact algorithm, that has successfully been used to generate expressive collision-, bouncing- and breaking-sounds, can also serve as the basis of a rolling-model.

To that end, the condition of contact must be enhanced to reflect the varying "offset" of the contact surface. Figure 9.9 gives a sketch of the process. The rolling object is here assumed to be locally spherical without "microscopic" surface details. These assumptions are unproblematic, since the micro details of the surface of the rolling object can be simply added to the second surface (to roll on) and the radius of the remaining "smoothed macroscopic" curve could be varied; in conjunction with following notions, even an assumed constant radius appears to be satisfactory for most modeling aims. It is important to observe that the contact between the two objects during the rolling is restricted to distinct points: the supporting surface is not fully "traced"/followed. The actual movement of the rolling object differs from this idealization due to inertia and elasticity. In fact, it is exactly the consequences of these physical properties, which are described by, and substantiate the use of, the dynamic impact model. To good approximation, the final vertical movement of the center of the ball is computed by use of the one-dimensional impact-model with the offset-curve shown in Figure 9.10.

Contact points and the resulting "substituting trajectory", that should ideally be applied to the one-dimensional impact model, are exemplified in Figure 9.10. The exact calculation of contact

[1] This fact is not reflected in the sketches, since here relative dimensions are highly unrealistic, namely exaggerated, for purposes of display.

Figure 9.10: Sketch of the effective offset-curve, resulting from the surface $s(x)$. The condition on the surface to be expressible as a function of x is clearly unproblematic in a “rolling” scenario.

points is computationally highly demanding: in each point x along the surface curve, i.e. for each sampling point in our practical discrete case (at audio rate), the following condition, which describes the momentary point of contact p_x, would need to be solved:

$$f_x(p_x) \overset{!}{=} max_{q \in [x-r, x+r]} f_x(q) \quad , \tag{9.6}$$

where

$$f_x(q) \triangleq s(q) + \sqrt{r^2 - (q-x)^2} \quad , \quad q \in [x-r, x+r] \quad . \tag{9.7}$$

The ideal curve would then be calculated from these contact points.

A simpler, computationally much cheaper strategy of “bridging” is sketched in Figure 9.11. This trajectory converges toward the ideal curve of Figure 9.10 for radii that are big as compared to the roughness of the surface. (Heuristic arguments also suggest an acoustic similarity of the two bridging strategies: both curves show similar characteristics of smoothed segments and occasional edges.)

In fact, in a first implementation sketch, even the extreme simplification of Figure 9.12, realized in a very simple algorithm, gave convincing results.

9.4.2 Surface

Different origins can be thought of, for the surface profile, which is a basis of the rolling-model developed in section 9.4.1. One possibility would be the scanning/sampling of real surfaces and use of such stored signals as input for the following stages of the model. This approach is sumptuous under the aspects of signal generation (a difficult scanning process) and memory and does not well

Figure 9.11: A simple method to approximate rolling-typical "bridging". The angle β has similar effect as a varying radius in the exact, ideal case.

Figure 9.12: "Bridging strategy realized by a very simple algorithm, that appeared practically useful nevertheless.

support the preliminaries of our modeling efforts: we are interested in expressive, flexible and effective sound cartoons rather than fixed realistic simulations of single specific scenes. Stored sound/signal files are generally hard to adapt to varying model attributes.

We thus prefer to use statistics-based "surface"-models, that can efficiently generate signals of varying attributes. It is common use in computer graphics to describe surfaces by fractal methods. One application of this idea to our one-dimensional case, the intersection curve through the surface along the path of rolling, leads to noise signals with a $1/f^{\beta}$ power spectrum; or equivalently, white noise filtered with this characteristic. The real parameter β here reflects the fractal dimension or roughness (see section 8.3.3).

Practical results of modeling following the so far developed methods became much more convincing when the bandwidth of the surface-signal was strongly limited. This does not surprise when one keeps in mind that typical surfaces of objects involved in rolling scenarios are generally smoothed to high degree. (In fact, it seems hard to imagine, what e.g. an uncut raw stone rolling on another surface, typically modeled as a fractal, let's say a small scale reproduction of the alps, would sound like?) Smoothing on a large scale, e.g. cutting and arranging pieces of stone for a stone floor, corresponds to high-pass-filtering, while smoothing on a microscopic level, e.g. polishing of stones, can approximately be seen as low-pass-filtering. In connection with this resulting band-pass, the $1/f^{\beta}$ characteristics of the initial noise signal loose in significance. We therefore opted for a very coarse approximation of this frequency curve by a second-order filter, whose steepness finally represents a "microscopic" degree of roughness. All frequencies in this low-level surface model have to vary proportional to a speed parameter; hereby, the amplitude of the surface-signal should be kept constant.

Of course, the parameters of the impact itself, in particular the elasticity constant k, must also be carefully adjusted to surface (e.g. material properties) and strongly contribute to the expressiveness of the model.

9.4.3 Higher-level features

Besides the "low-level" parameters motivated and described in the last section (9.4.1) typical rolling scenarios show characteristic "macroscopic" features, that strongly contribute to the acoustic perception, and can not be described by the means developed so far. Many "ground" surfaces contain typical patterns of more or less regular nature that are not suitably characterized as filtered fractal noise. Everyday experience can easily testify more or less periodic (accordingly) auditory features of rolling-sounds arising from such surface patterns. Prominent examples are the regular joints of stone- or wooden floors, the periodic textures of textiles or the pseudo-periodic furrows in most wooden boards. Single "dips" or outstanding irregularities on the surface of a rolling object can be grouped in the same category, since they are "recalled" periodically in the turning movement. Such features can be modeled, e.g. with pulse-like signals of constant or slightly changing frequency; a (e.g.) polynomial or sinusoidal approximation may here be useful, with a "smoothness" parameter related to the degree of the approximating function, accounting for the "edginess" of the pattern to be modeled. Again all frequencies must vary proportional to a speed parameter.

Another macroscopic observation has to be accounted for as relevant for rolling-sounds: For rolling objects that are not perfectly spherical (in the section relevant for the movement), the velocity of the point of contact on both surfaces and the effective force pressing the rolling object to the ground vary periodically. In order to model such deviations from perfect sphericity, these two parameters must be modulated; a good choice are obviously sinusoidal or other narrow-band modulation signals. Objects that differ too much from a spherical shape, that are too edgy, do not roll…

Finally it is to be noted that, like in everyday listening, acoustic rolling scenarios are recognized and accepted more easily with "typical dynamics": as an example, consider the sound of a falling marble ball, that bounces until constant contact to the ground is reached, now the rolling action gets acoustically clear and the average speed slowly declines to zero.

9.5 Crumpling

For what we have seen in chapter 8 the impact model yields, at its interface level, several controls which are highly useful for recreating realistic collisions between objects made of different materials. Meanwhile, the same controls enable the user to interact with those virtual objects directly as well.

As long as ecological sounds describe higher-level events such as crushing or walking [54], then we need representations which are more articulate than the basic impact model. Due to physics-based control we can follow a bottom-up approach, and start from the elementary model to build up higher-level models.

In the following of this section we will describe the way individual impacts can be assembled together to form a single crushing sound. Crushing sounds are the result of a statistical rather than deterministic sequence of impacts. This means that we are not asked to look for (deterministic) closed-form formulas valid in the discrete time, expressing relationships between different types of sounding objects—those formulas are required in the study of contact sounds [9]. We instead will find a way to create consistent (from a psychophysical point of view) collections of "atomic" (impact) sounds starting from the stochastic description of so-called *crumpling sounds*.

9.5.1 Crumpling sound

Crumpling sound occurs whenever a source emission can be modeled as a superposition of crumpling events.

Aluminum cans emit a characteristic sound when they are crushed by a human foot that, for example, compresses them along the main axis of the cylinder. This sound is the result of a composition of single crumpling events, each of those occurring when, after the limit of bending resistance, one piece of the surface forming the cylinder splits into two facets as a consequence of the force applied to the can.

The exact nature of a single crumpling event depends on the local conditions the surface is subjected to when folding occurs between two facets. In particular, the types of vibrations that are produced are influenced by shape, area, and neighborhood of each facet. Moreover, other factors play a role during the generation of the sound, such as volume and shape of the can. The can, in its turn, acts as a volume-varying resonator during the crumpling process.

A precise assessment of all the physical factors determining the sound which is produced by a single crumpling event is beyond the scope of this work. Moreover, there are not many studies available in the literature outlining a consistent physical background for this kind of problems. On the other hand it is likely that our hearing system cannot distinguish such factors, but the most relevant ones. For this reason we generate individual crumpling sounds using the impact model.

Studies conducted on the acoustic emission from wrapping sheets of paper [121] concluded that crumpling events do not determine *avalanches* [216], so that fractal models in principle cannot be used to synthesize crumpling sounds [214]. Nevertheless, crumpling paper emits sounds in the form of a stationary process made of single impulses, whose individual energy E can be described by the following power law:

$$P(E) = E^{\gamma} \quad , \tag{9.8}$$

where γ has been experimentally determined to be in between -1.3 and -1.6. On the other hand a precise dynamic range of the impulses is not given, although the energy decay of each single impulse has been found to be exponential.

Another (perhaps the most important) factor determining the perceptual nature of the crumpling process resides in the temporal patterns defined by the events. A wide class of stationary temporal sequences can be modeled by *Poisson's processes*. According to them, each time gap τ between two subsequent events in a temporal process is described by an exponential random variable with *density* $\lambda > 0$ [192]:

$$P(\tau) = \lambda e^{\lambda\tau} \text{ with } \tau \geq 0 \quad . \tag{9.9}$$

Assuming a time step equal to T, then we simply map the time gap over a value kT defined in the discrete-time domain:

$$k = \text{round}(\tau/T) \quad , \tag{9.10}$$

where the operator round gives the integer which is closest to its argument value.

The crumpling process consumes energy during its evolution. This energy is provided by the agent that crushes the can. The process terminates when the transfer of energy does not take place any longer, i.e., when a *reference energy*, E_{tot}, has been spent independently by each one of the impulses forming the event s_{tot}:

$$s_{\text{tot}}[nT] = \sum_i E_i s[nT - k_i T] \quad \text{with} \quad E_{\text{tot}} = \sum_i E_i \quad , \tag{9.11}$$

where $s(nT)$ is a signal that has unitary energy, accounting for each single crumpling.

At this point, the individual components of the process and their dynamic and temporal statistics have been decided. Yet, the dynamic range must be determined.

Suppose to constrain E to assume values in the range $[m, M]$. The probability P that an individual impulse falls in that range is, using the power law expressed by (9.8):

$$P[m \leq E < M] = \int_m^M E^\gamma dE = 1 \quad . \tag{9.12}$$

This equation allows to calculate an explicit value for m if we set M to be, for example, the value corresponding to full-scale, beyond which the signal would clip. In this case we find out the minimum value coping with (9.12):

$$m = \{M^{\gamma+1} - \gamma - 1\}^{\frac{1}{\gamma+1}} \quad . \tag{9.13}$$

9.5.2 Driving the impact model

We still have to determine rules for selecting the driving parameters of the impact model each time a crumpling sound is triggered.

During the crushing action over an object, creases become more and more dense over the object's surface. Hence, vibrations over the facets increase in pitch since they are bounded within areas that become progressively smaller. This hypothesis inspires the model that is used here to determine the pitch of an individual crumpling sound.

Figure 9.13: Sketch of the procedure used to calculate the pitch of the impulses as long as the process evolves.

Let us consider a segment having a nominal length D_0, initially marked at the two ends. Let us start the following procedure: Each time a new impulse is triggered, a point of this segment is randomly selected and marked. Then, two distances are measured between the position of this mark and its nearest (previously) marked points. The procedure is sketched in Figure 9.13, and it is repeated as long as some energy, as expressed by (9.11), is left to the process.

The values L_i and R_i, corresponding to the distances calculated between the new mark m_i (occurring at time step k_i) and the leftward and rightward nearest marks (occurring at previous time steps), respectively, are used as absolute values for the calculation of two driving frequencies, f_L and f_R, and two decay times, τ_L and τ_R, and also as relative weights for sharing the energy E_i between the two impacts, x_{f_L,τ_L} and x_{f_R,τ_R}, forming each crumpling sound:

$$E_i s[nT - k_iT] = E_i \frac{L_i}{L_i + R_i} x_{f_L,\tau_L}[nT - k_iT] + E_i \frac{R_i}{L_i + R_i} x_{f_R,\tau_R}[nT - k_iT] \quad , \tag{9.14}$$

where the driving frequencies (decay times) are in between two extreme values, f_{MAX} (τ_{MAX}) and f_{MIN} (τ_{MIN}), corresponding to the driving frequencies (decay times) selected for a full and a minimum portion of the segment, respectively:

$$\begin{aligned} f_L &= f_{\text{MAX}} - \frac{L_i}{D_0}(f_{\text{MAX}} - f_{\text{MIN}}) \\ f_R &= f_{\text{MAX}} - \frac{R_i}{D_0}(f_{\text{MAX}} - f_{\text{MIN}}) \quad , \end{aligned}$$

in which the symbols f must be substituted by τ in the case of decay times.

We decided to operate on the so-called “frequency factor” and decay time of the impact model: the former is related to the size of the colliding object; the latter accounts for the object material

Figure 9.14: Sonogram of the prototype sound of a crushing can.

[9]. For what we have said, we considered both of them to be related to the crumpling facet area: the smaller the facet, the higher-pitched the fundamental frequency and the shorter the decay time of the emitted sound.

9.5.3 Can crushing

Crushing occurs in consequence of some force acting on the can. This action is usually performed by an agent having approximately the same size as the can surface, such as the sole of a shoe.

As the agent compresses the can, sound emission to the surrounding environment changes since the active emitting surface of the can is shrinking, and some of the creases become open fractures in the surface. Moreover, we suppose that the internal pressure in the can is maximum in the beginning of the crushing process, then relaxes to the atmospheric value as long as the process evolves, due to pressure leaks from the holes appearing in the surface, and due to the decreasing velocity of the crushing action. Those processes, if any[2], have a clear effect on the evolution in time of the spectral energy: high frequencies are gradually spoiled of their spectral content, as it can be easily seen from Figure 9.14 where the sonogram of a real can during crushing has been plotted.

The whole process is interpreted in our model as a time-varying resonating effect, simply realized in our model through the use of a low-selectivity linear filter whose frequency cut s_{tot} is slid

[2] We are still looking for a thorough explanation of what happens during crushing.

toward the low-frequency as long as the process evolves.

Lowpass filtering is performed using a first-order lowpass filter [177]. In the case of crushing cans we adopted the following filter parameters:

- lowest cutoff frequency $\Omega_{\text{MIN}} = 500$ Hz
- highest cutoff frequency $\Omega_{\text{MAX}} = 1400$ Hz.

Using those parameters, the cutoff frequency is slid toward the lowest value as long as energy is spent by the process. More precisely, the cut frequency ω_i at time step k_i is calculated according to the following rule:

$$\omega_i = \Omega_{\text{MIN}} + \frac{E_{\text{tot}} - \sum_{k=1}^{i} E_k}{E_{\text{tot}}} \left(\Omega_{\text{MAX}} - \Omega_{\text{MIN}}\right) \quad . \tag{9.15}$$

This kind of post-processing contributes to give a smooth, progressively "closing" characteristic to the crumpling sound.

Parameterization

Several parameter configurations have been tested during the tuning of the model. It has been noticed that some of the parameters outlined above have a clear (although informal) *direct* interpretation:

- E_{tot} can be seen as an "image" of the *size*, i.e., the height of the cylinder forming the can. This sounds quite obvious, since E_{tot} governs the time length of the process, and this length can be in turn reconducted to the can size. Sizes which are compatible with a natural duration of the process correspond to potential energies ranging between 0.001 and 0.1;
- low absolute values of γ result in more regular realizations of the exponential random variable, whereas high absolute values of the exponential statistically produce more peaks in the event dynamics. Hence, γ can be seen as a control of *force* acting over the can. This means that for values around -1.5 the can seems to be heavily crushed, whereas values around -1.15 evoke a softer crushing action. Thus, γ has been set to range between -1.15 and -1.5;
- "soft" alloys forming the can can be bent more easily than stiff alloys: Holding the same crushing force, a can made of soft, bendable material should shrink in fewer seconds. For this reason, the parameter p_s governing the frequency of the impulses in the Poisson process can be related to the material stiffness: the higher p_s, the softer the material. *Softness* has been set to range between 0.001 (stiff can) and 0.05 (soft can).

We noticed that variations in the definition and parameterization of the crumpling sound s can lead to major differences in the final sound. On the other hand the statistical laws which we used for generating crushing sounds are general enough for reproducing a wide class of events, including, for example, paper crumpling and plastic bottle crushing. In that case s must be properly shaped to accommodate for different kinds of events.

Figure 9.15: Screenshot of the pd module implementing the crushing can model.

Implementation as pd patch

Crushing cans have been finally implemented as a pd patch [197]. pd allows a modular implementation: hence, we could maintain the crumpling sound synthesis model, the higher-lever statistical module and the time-varying post-processing lowpass filter decoupled inside the same patch (see Figure 9.15). The only limitation with this implementation is represented by the firing rate with which the statistical module (labeled as control_crump) can feed control data to the two sound synthesis modules (labeled as sound_synth˜) producing the signals x_{f_L} and x_{f_R}, respectively. This limitation comes from the presence of a structure containing the statistical module in loop with a delay block, which limits the shortest firing rate to 1 ms.

On the other hand the chosen implementation allows a totally independent design of the statistical and the sound synthesis modules. More precisely, the former has been realized in C language, as a pd *class*, whereas the latter has been implemented as a sub-patch nesting inside it several previously existing building blocks, which come together with pd.

This modular/nested approach leaves the patch open to independent changes inside the modules, and qualifies the patch in its turn for use as an individual block inside higher-level patches.

For this reason we could straightforwardly integrate crushing sounds in a framework including specific rules for producing walking and running sounds.

9.6 Some audio cartoons of fluid movements

The cartoonifications of fluid sounds described in the following can give a good example how effective sound models can be constructed even on the basis of very simple techniques of signal generation. Elementary sounds that are far from any ecological association can gain a strong meaning or expressive potential when distributed and temporally grouped appropriately. In this current case, primitive, rather abstract low-level "bursting-bubble" events are combined in characteristic patterns to dropping- and streaming-models. The streaming-model is further completed with a filter stage to depict a filling-/emptying-process, supplying e.g. a useful sonification for a download-process.

Realism is not the final aspired and achieved result of the design process. Instead, some acoustic attributes are exaggerated, e.g. the characteristic resonance behavior of a filling/emptying vessel, which is not easily perceived from a real sound. Also, in contrast to recorded sound files, the parameters of the objects can be flexibly adjusted, e.g. the speed of filling, the initial/final fluid level, the sharpness of the vessel resonance or the viscosity and the amount of gas in the fluid.

9.6.1 Audio kernel

The basic audio algorithm is based on a rough examination of typical movements of the surface of a fluid [97]: an upcoming bursting bubble in water forms a hollow resonator that first opens up with an impulse when the bubble reaches the surface and successively gets "flattened out". During that process the resonance behavior changes accordingly; the frequency of the dominant resonant peak changes with the depth of the pit and the "sharpness" of the resonance is related to its flatness. An object falling into a fluid can initially raise a similar behavior with opposite direction of movement. This process starts with an impulse at the first contact, and other impulses occur when emerging pits close. Consequently inclosed amounts of gas finally rise up in bubbles again.

These qualitative examinations are exploited in the lowest-level sound object, which uses a pure sine wave with a frequency-envelope tuned according to the portrayed resonance development. An amplitude-envelope creates initial and terminating impulses and accounts for the decaying peak levels. Not surprisingly, this simple first building block produces sounds that are highly artificial and not realistic at all. Main acoustic characteristics however prove to be sufficiently represented for convincing higher sound objects. (It is worth to notice, that people tend to instinctively create a similar sound when trying to vocally imitate a bursting bubble or related fluid movements.) Envelope parameters are mapped on model attributes, such as e.g. the size of a bubble, with a suitable random element. Figure 9.16 shows a pd implementation of the afore-described synthesis procedure.

Figure 9.16: `pd` patch using a sine oscillator and envelopes to model an elementary fluid movement, like the bursting of an isolated bubble.

9.6.2 Fluid scenarios

Dropping/falling into fluid

The fine-scale fluid motions depicted in section 9.6.1 almost never occur in isolation. Typical every-day fluid scenarios involve a huge number of such micro-events, in a typical distribution.

An object falling into water firstly triggers many small drops to be emitted and fall down again. Another bigger drop typically jumps up at the closing of an emerged pit and after the disappearance of an upcoming bubble. So, each falling drop of sufficient size in turn causes following smaller ones. This recursive process was successfully modeled with the help of a lisp program in combination with a CSOUND realization of the sound object developed in section 9.6.1. During the implementation of an according `pd` object, the graphical programming environment proved less well adapted to the realization of such higher-level control structures.

Figure 9.17: Maisy drinking with a straw (From “Happy Birthday, Maisy” by Lucy Cousins, Candlewick Press, 1998.)

Streaming

Continuous fluid movements, such as those caused by a constant stream or a filling action, in contrast to a dropping-event, acoustically seem to be rather characterized by a sustained layering of micro-events of the form of section 9.6.1. The distribution of the parameters of these “micro-bubbles” and their temporal density appear to be cues for the viscosity of the fluid and other characteristics of a scenario, such as fluid volume or streaming velocity.

A “streaming” `pd` object uses an intuitively chosen equal distribution of the lower-level blocks of section 9.6.1. Another realization on the basis of CSOUND and lisp, with a two-band distribution gives a more convincing result. Both implementations are combined with a filtering stage, modeling the basic resonance of e.g. a bottle, that again changes with the remaining volume of air. A parametric equalizer as obvious simple choice completes a convincing audio-cartoon of a filling-process.

As an example for a more abstract use of our sound objects, serves a sonification of a scene from children’s book. The combination of the streaming-object with a comb-filter could be interactively controlled simultaneously with a simple adjustable cardboard figure in the book: an “aristocratic mouse”, drinking lemonade with a straw! In the book, the reader can regulate the fluid level of the mouse’s drinking glass (see Figure 9.17).

That control might be coupled with the delay time (i.e. base frequency) of the comb filter (see Figure 9.18). Together with typical envelopes for a starting/finishing streaming-action an additional inspiring acoustic feedback is given. Of course, drinking with a straw generally does not

Figure 9.18: The two highest levels of the “drinking mouse” `pd` patch, that triggers the audio-atom of Figure 9.16 in an appropriate pattern and uses comb filters to express the progressing drinking action.

generate a sound close to this one, if any remarkable noise at all. Nevertheless, the free association of this moving liquid-like sound, with that cartoon scene, seems natural. Asked to spontaneously sonify this scene with their voice, many people would probably produce similar noises, although it does not really sound like having-a-cocktail (hopefully).

Chapter 10

Synthesis of distance cues: modeling and validation

Federico Fontana, Laura Ottaviani and Davide Rocchesso
Università di Verona – Department of Computer Science
Verona, Italy
fontana@sci.univr.it, ottaviani@sci.univr.it, davide.rocchesso@univr.it

10.1 Introduction

Previous research in distance perception has mostly investigated the human capability to evaluate how far a sound source is from ourselves. Psychophysical experimentation involving tasks of distance recognition raises several issues, both methodological and technical. The interpretation and analysis of the subjects' impressions, when sounds are played away from the near-field, is complicate to perform as well.

It is probably for the reasons above that, even today, most of the systems aiming to add distance attributes to sounds either rely on reverberation models that were designed for other, more general purposes, or reproduce simplified physical contexts in which sounds are provided with elementary information about distance.

Our approach to distance rendering by synthesis of virtual spaces pursues an *objective* goal. In fact, we propose a model that certainly reflects a physical rather than perceptual approach. In addition, we exploit the inherent versatility of this model to explore physical environments that, independently of their consistency with the everyday experience of the surrounding space, provide to a listener cues that are capable of evoking a well-defined sense of distance.

Psychophysical experiments, conducted through both headphone and loudspeaker listening tests, will confirm the suitability of this approach.

Humans use several senses simultaneously to explore and experience the environment. Unfortunately, technological limitations often prevent computer-based systems from providing genuine multi-sensory channel (that is, *multimodal*) displays. Fortunately, the redundancy of our sensory system can be exploited in order to choose, depending on cost and practical constraints, the display that is the most convenient for a given application.

Providing access to information by means of audio signals played through headphones or loudspeakers is very attractive, especially because they can elicit a high sense of engagement with inexpensive hardware peripherals. Namely, one may be tempted to transfer spatial information from the visual to the auditory channel, with the expected benefits of enlarging the informational flow and lowering the load for the visual channel. However, we should bear in mind that vision and hearing play different roles in human perception. In particular, space is not considered to be an "indispensable attribute" of perceived sounds [140].

In audio-visual displays, the effectiveness of communication can be maximized if the visual channel is mainly devoted to spatial (and environmental) information, while the auditory channel is mainly devoted to temporal (and event-based) information. However, there are several situations where the spatial attributes of sound become crucial:

1. in auditory warnings, where sounds are used to steer the visual attention;
2. where it is important to perceive events produced by objects that are visually occluded or out of the visual angle;
3. for visually impaired users, where visual information is insufficient or absent.

Furthermore, if the "soundscape" is particularly rich, the spatial dislocation of sound sources certainly helps the tasks of separation and understanding of events, streams, and textures.

Much research has been dedicated to spatial auditory displays, with special emphasis on directional cues [205], but the literature on the perception and synthesis of the range of sources is quite limited [157, 257, 169]. Moreover, part of this literature does not provide clear specifications about the experimental conditions, particularly concerning listener's impressions about the auditory event and the *blur* experienced during the recognition task [21].

An interesting approach to distance evaluation [157] compares the auditory to the visual distance. In some experiments the distance evaluation were reported by the listeners verbally, whereas under other conditions the subjects had to walk toward the sound or visual source.

The experiments in [257] have different goals: to investigate the cues in distance localization, and to understand how they are combined for stable estimations. It results demonstrated that the subjects underestimate far-field distances, while they overestimate near-field distances. Moreover, a factor called *specific distance tendency* is introduced to describe the bias toward a specific perceived distance when all distance cues are removed. The specific distance tendency is estimated to be approximately 1.5 m.

Our purpose is to analyze only *static distance cues*, since in our experiments participants performed the distance estimation from a fixed location in space. What most psychoacoustic studies have found is that we underestimate significantly the distance of sources that are located farther than a couple of meters from the subject.

The acoustic cues accounting for distance are mainly *monaural*:

1. *Spectrum* conveys distance information as well, if the listener has enough familiarity with the original sound. In that case, spectral changes introduced in the direct signal by air loss and/or sound reflection over non-ideal surfaces can be detected by the listener, and hence converted into distance information [21].

2. *Intensity* plays a major role, especially with familiar sounds in open space. Theoretically, intensity in the open space decreases by 6 dB for each doubling of the distance between source and listener [181].

3. *Direct-to-reverberant energy ratio* affects perception in closed spaces or reverberant outdoor environments. The reverberant energy comes from subsequent reflections of the direct sound, each of them having amplitude and time delay that vary with the characteristics of the enclosure, and with the source and listener's positions.

Also, the existence of *binaural* cues has been demonstrated. These cues are particularly important in the case of nearby sources [40]. Monaural cues coming from nearby sources have been hypothesized as well [21].

The first cue can be exploited to display very large distances, because the spectral cues are relevant only for long paths. The second cue is not very useful in auditory displays and sonification, because it imposes restrictions to the listening level and it may lead to uncomfortable soundscapes.

The third cue can be exploited to synthesize spatial auditory displays of virtual sources in the range of about ten meters. This cue is inherently connected with spatial hearing inside enclosures: in this listening context, a certain amount of reverberant energy is conveyed to the listener. According to Guski [108], the effect of reflecting surfaces shows an overall increment in localization accuracy in the case when a sound-reflecting surface is placed on the floor. On the contrary, the percentage of correct estimation decreases with a sound-reflecting surface put on the ceiling, whereas a surface located on one side of the room doesn't affect the performance. A sound-reflecting surface on the floor increases both the direction and height localization accuracy.

The virtual reproduction of wall reflection is traditionally provided by artificial reverberators, which add reverberant energy to sounds.

The effects of the characteristics of the environment on sound are difficult to model and highly context-dependent. Moreover, the psychophysical process that maps the acoustics of a reverberant enclosure to the listener's impressions on that enclosure is still partially unknown [17]. For this reason, artificial reverberators are typically the result of a *perceptual* design approach [90], which had the fundamental advantage of leading to affordable architectures working in real-time, and resulted in several state-of-the art realizations, providing high-quality rendering of reverberant environments [128]. Nevertheless, most of these realization do not deal with distance rendering of the sound source.

On the contrary, the *structural* design philosophy focuses on models whose properties have a direct counterpart in the structural properties that must be rendered, such as the geometry of an enclosure or the materials the wall surfaces are made of. Thanks to that approach, their driving parameters translate into corresponding model of behavior directly. Unfortunately, structural models resulted to be either too resource-consuming, or, when simplified to accommodate the hardware requirements (i.e., the real-time constraint), excessively poor in the quality of the audio results.

The emerging *auditory display* field shifts the focus on the usability of the auditory interface rather than on the audio quality *per se*. For the purpose of enhancing the effectiveness of display, it is often useful to exaggerate some aspects of synthetic sounds. In spatial audio, this led to systems for supernormal auditory localization [67]. In this work, we will exaggerate certain characteristics of sound to improve the localization of a virtual sound source.

As some recent literature pointed out, presenting distance cues to a listener in an auditory display does not necessarily require an individual optimization of the display. In fact, non-individualized Head-Related Transfer Functions influence the quality of those cues only to a small extent [258]. This enables us to focus on the objective mechanisms of distance rendering, neglecting the matter of subjective tuning.

We use the structural approach to reverberation to design a virtual resonator that enhances our perception of distance. For example, let us consider a child playing inside one of those tubes that we find in kindergartens. If we listen to the child by staying at one edge of the tube, we have the feeling that he/she is located somewhere within the tube, but the apparent position turns out to be heavily affected by the acoustics of the tube. Using a virtual acoustic tool, we experimented several tube sizes and configurations, until we found a virtual tube that seems to be effective for distance rendering. In a *personal* auditory display [169], where the user wears headphones and hears virtual as well as actual sound sources, these tubes will be oriented in space by means of conventional 3D audio techniques [205], so that the virtual sound sources may be thought to be embedded within virtual acoustic beams departing from the user's head.

10.2 Acoustics inside a tube

The listening environment we will consider is the interior of a square-section cavity having the aspect of a long tube, sized 9.5×0.45×0.45 meters. The internal surfaces of the tube are modeled to exhibit natural absorption properties against the incident sound pressure waves. The surfaces located at the two edges are modeled to behave as *total* absorbers (see Figure 10.1) [143].

A careful investigation on the physical properties of this acoustic system is beyond the scope of this work. The resonating properties of cavities having similar geometrical and absorbing properties (for example organ pipes) have been previously investigated by researchers in acoustics [76].

It seems reasonable to think that, although quite artificial, this listening context conveys sounds that acquire noticeable spatial cues during their path from the source to the listener. Given the peculiar geometry of the resonator, these cues should mainly account for distance. The edge surfaces have been set to be totally absorbent in order to avoid echoes originating from subsequent reflections of the wavefronts along the main direction of wave propagation. In fact, these echoes would be ineffective for distance recognition in the range specified by the tube size, and would also originate a side-effect annoying the listener.

The resonating environment is structurally simple although capable of dealing with an interest-

Figure 10.1: The listening environment. All sizes in meters.

ing range of the physical quantity to be rendered. In particular, it exhibits a convenient trade-off between the distance range and the resonator volume (the larger the volume, the longest the computation time).

In this environment, we put a sound source at one end of the tube (labeled with S in Figure 10.1) along the main axis. Starting from the other end, we move a listening point along 10 positions $x_{10}, \ldots, x_1$ over the main axis, in such a way that, for each step, the source/listener distance is reduced by a factor $\sqrt{2}$. The resulting set of distances expressed in meters, X, is shown in Figure 10.1:

$$\begin{aligned} X &= \{x_1, \ldots, x_{10}\} \\ &= \{0.42, 0.59, 0.84, 1.19, 1.68, 2.37, 3.36, 4.75, 6.71, 9.5\} \quad . \end{aligned} \tag{10.1}$$

An obvious question arises prior to any investigation on distance rendering: why not render distance in an auditory display by simply changing the loudness of sounds as a function of their distance from the listener? The answer is twofold:

- distance recognition by intensity is effective only if the sound source is familiar [21]. Conversely, a resonating environment, once become familiar to the listener, adds unique "fingerprints" to the sound emitted by the source, so that the listener has more chances to perform a recognition of distance also in the case of unfamiliar sounds;

- intensity in open space follows a 6 dB law for each doubling of distance. This means that a wide dynamic range is required for recreating interesting distance ranges in virtual simulations of open spaces. This requirement, apart from inherent technical complications due to hardware constraints, might conflict with the user's need of hearing other (possibly loud) events in the display. These events would easily mask farther virtual sound sources, especially in the case when the auditory display is designed to work with *open* headphones or

Figure 10.2: Particular of a volume section. The lossless scattering junction in the center is connected to other junctions via waveguides 2, 3, 4, and 6. Waveguide 1 leads to a totally absorbing section of wall. Waveguide 5 leads to a partially absorbing section of wall, modeled using a waveguide filter. The filled triangles represent oriented unit delays.

> *ear-phones* [141]. In this case, dynamic compression of sounds can be taken into account, even if this would possibly lead to a corresponding reduction of the range perceived by the listener.

Summarizing, the proposed environment should lead to a robust simulation also with respect to unfamiliar sounds, and to a broad perceived range obtained by a compressed loudness-to-distance law.

10.3 Modeling the listening environment

The square tube has been modeled by means of finite-difference schemes. Such schemes provide a discrete-space and time formulation of the fundamental partial differential equation accounting for 3D wave propagation of pressure waves along an ideal medium [229]. In this way, they clearly devise a structural approach to the problem of modeling a reverberant environment. Their simplicity is a key feature for this research, that came useful especially during the preliminary informal listening of several tubular environments differing in size, shape and absorption properties.

In particular, a *wave*-based formulation of the finite-difference scheme has been used, known as the Waveguide Mesh, that makes use of the wave decomposition of a pressure signal p into its wave components p^+ and p^-, such that $p = p^+ + p^-$ [69]. By adopting this formulation the spatial domain is discretized in space into equal cubic volume sections, and each of them is modeled as a lossless junction of ideal waveguides, scattering 6 input wave pressure signals coming from orthogonal directions, $p_1^+, \ldots, p_6^+$, into corresponding output waves, $p_1^-, \ldots, p_6^-$, going to opposite directions (see Figure 10.2).

It can be shown that, given the length d_W of each waveguide, pressure waves travel along the Waveguide Mesh at a speed equal to

$$c_W \leq \frac{1}{\sqrt{3}} d_W F_s \quad , \tag{10.2}$$

where F_s is the sampling frequency, and the symbol $\leq$ means that some spatial frequencies travel slower along the mesh. This phenomenon is called *dispersion* [229], and its main effect is a detuning of high frequencies. Dispersion is not considered to be important for the application.

Assuming the velocity of sound in air equal to 343 m/s, and setting $F_s = 8$ kHz, we have from (10.2) that each waveguide is about 74.3 mm long. Thus, the mesh needs $127 \times 5 \times 5 = 3175$ scattering nodes to model our tube. The reader should note that the sampling frequency has important effects on the computational requirement of the model. Our choice is oriented to efficiency rather than sound realism: reliable distance cues should be conveyed also using lower sampling frequencies.

The Waveguide Mesh has already been used in the simulation of reverberant enclosures [211]. For our purpose, it allows to deal with the boundary of the propagation domain quite effectively [68]. In fact, it enables the direct application of *waveguide filters* at the mesh boundary, to model the reflection properties of the internal surfaces of the tube [123].

More in detail, each waveguide branch falling beyond the boundary of the tube is terminated with a spring/damper system, that models a simplified resonating/absorption property of the surface at the waveguide termination. This system is algebraically rearranged into a Waveguide filter, then discretized into a Waveguide Digital filter establishing a transfer function, $B(z)$, between pressure waves going out from, and incoming to the closest scattering junction:

$$B(z) = \frac{P_i^+(z)}{P_i^-(z)} \quad . \tag{10.3}$$

For example, it is $i = 5$ in Figure 10.2. Using this physical system, the resulting filter terminations are made of 1st-order pole-zero filters. Despite this simplicity, these filters can be tuned to simulate the reflecting properties of real absorbing walls with good precision [143].

Considering that the surfaces at the two terminations of the tube have been set to be totally absorbing (this meaning that $p^+ \equiv 0$), the total number of boundary filters is $127 \times 5 \times 4 = 2540$.

10.4 Model performance

Measures conducted on tube impulse responses are summarized in Figure 10.3 and Figure 10.4. Figure 10.3 shows spectral differences existing between sounds auditioned close to and far from the sound source, for both the left and right channel. Figure 10.4 shows how the signals average magnitudes, defined by the value $10 \log \frac{1}{n} \sum_n [s(n)]^2$ where s is the signal to be studied, and referenced to 0 in correspondence of the the closest position, vary with distance: These variations

Figure 10.3: Magnitude spectra of signals picked up at distances of 1 m and 10 m with a left and right pickup, ordered by decreasing magnitude.

show that a dynamic range smaller than the 6 dB law is needed for rendering distances using the proposed method.

In particular, Figure 10.4 shows that the right-channel magnitudes diverge from the left ones, as long as the range becomes greater than about 1 m. This divergence does not appear in reverberant sounds taken from real-world environments. This effect can be certainly reconducted to the peculiarity of the listening environment proposed here. Nevertheless, a careful analysis of the side-effects coming from using a coarse-grained realization of the Waveguide Mesh as a model of the listening environment should be carried out, to assess the precision of the plots depicted in Figure 10.4.

10.5 Psychophysical validation

We conducted two experiments in two different listening conditions: one using headphones and one using loudspeakers, in order to evaluate the model response in the two reproduction situations. There is a branch of auditory display that studies the differences existing between headphone and loudspeaker presentation of spatialized sounds [129]. In our model we have not added any specific adaptation to different devices. However, we want to investigate the two cases.

The tests apply the magnitude estimation method. We investigated how users scaled the perceived distance and, hence, whether our model is effective or not.

Figure 10.4: Averaged magnitudes of all acquired signals as a function of distance, together with magnitude values in the ideal open-space case. Above: right channel. Below: left channel. Middle: average between left and right channels.

10.5.1 Listening through headphones

The experiment, in the headphone reproduction condition, involved 12 users (4 females and 8 males), with age between 22 and 40, who voluntarily participated to the experiment. They studied or worked at the University of Verona. All of them were naive listeners and nobody referred to have hearing problems.

Stimuli

The sound set was synthesized by putting a sound source at one end of the virtual tube, along the main axis, and acquiring ten stereophonic impulse responses along positions $x_{10}, \ldots, x_1$ (see Figure 10.1 and Eq. (10.1)).

The right channel of the stereophonic sound accounts for acquisition points exactly standing on the main axis, whereas the left channel accounts for points displaced two junctions far from that axis, this corresponding to an interaural distance of about 15 cm.

The impulse responses obtained in this way have been convolved with a short, anechoic sample of a cowbell.

Procedure

The experiment was conducted in a quite, but not acoustically isolated room and the setup involved a PC Pentium III, with a Creative SoundBlaster Live! soundcard. During this first experiment, sounds were played through Beyerdynamic DT 770 closed headphones.

The users listened to a sequence of 30 sounds, consisting of the 10 sounds in the stimuli set randomly repeated for three times, and they were asked to estimate the perceived distance from the sound source, using headphones.

The participants, without training, had to rate each distance with a value in meters (either integer or decimal), starting from the first one, and associating a value to the other ones, proportionally to the first estimation. The collected values defined scales that depended on the individual listeners' judgments. These scales ranged from 0.2-8 meters (user no. 8) to 1-30 meters (user no. 5).

The three judgments given for each sound were then geometrically averaged for each participant, and the resulting values were used to calculate a mean average. Subtracting it from the individual averages, we adjusted the listeners' judgments to obtain a common logarithmic reference scaling [70].

Results and Observations

In Figure 10.5 the distance evaluations for each listener are shown as functions of the source distance, together with the corresponding linear functions obtained by linear regression. The average slope is 0.61 (standard deviation 0.21), while the average intercept is 0.46 (standard deviation 0.21).

In Figure 10.6 the perceived distance averaged across subjects is plotted as function of the source distance, together with the relative regression line ($r^2 = 0.76$, $F(1, 8) = 25.84$, $p < 0.01$).

From this first experiment we observed that users overestimate the distance of close sound sources, and that they reduce this overestimation for greater distances. This result is interesting since it partially contradicts Zahorik [257], who worked with real sounds for the tests, and reported the tendency of listeners to overestimate short distances, and underestimate long distances. In our model, the point of correct estimation is more distant compared with Zahorik. More precisely, our regression line has an offset upward, in a way that the point of correct estimation moves toward longer distances. This result can be interpreted as a consequence of the exaggerated reverberant energy produced by our model.

10.5.2 Listening through loudspeakers

The second experiment involved 10 volunteers (4 females and 6 males), 4 of which participated also to the first experiment. They worked or studied in our department and they were aged between 23 and 32. All the subjects reported to have normal listening.

Stimuli and Procedure

The stimuli were the same of the previous test, and the experiment was conducted in the same quite room, but the reproduction system, in this second experiment, used loudspeakers. The participants sat at a distance of 1.5 m from a pair of Genelec 2029B stereo loudspeakers, 1 m far from each other, and a Genelec subwoofer was located in between the loudspeakers.

Figure 10.5: Headphone listening: Individual distance evaluations together with individual linear regression lines. a: intercept. b: slope.

The users were blindfolded, in order to minimize the influence of factors external to the experiment. They were asked to evaluate metrically the distance of the sound source from the listening point communicating its value to the experimenter. The first value, as in the previous test, determined the subjective scale.

Results and Observations

In Figure 10.7 we report, for each participant, the distance evaluations as functions of the source-listener distance, together with the corresponding linear functions obtained by linear regression. The average slope is 0.53 (standard deviation 0.17), while the average intercept is 0.50 (standard deviation 0.36).

In Figure 10.8 the perceived distance averaged across subjects is plotted as function of the source-listener distance, together with the relative regression line ($r^2 = 0.85$, $F(1, 8) = 45.76$, $p < 0.01$).

The loudspeaker test led to results that are similar compared with the headphone test results. In fact it is evident that, in both cases, there is a distance overestimation for closer sound sources,

Figure 10.6: Average distance evaluation together with linear regression line. a: intercept. b: slope.

that reduces as long as the distance increases.

There is only one user (no. 10) whose individual scale ranged between 0.1-2 meters, and who perceived all the sound sources to be closer compared with the other listeners.

10.6 Conclusion

A comparison between the two experiments gives interesting hints. The users' responses are similar in both the reproduction conditions. Although our model does not deal with the issue of sound reproduction, it fits both reproduction systems [129, 258].

Moreover, there is an exaggeration especially in rendering close sound sources, probably due to the amount of reverberant energy existing in that case. The point of correct estimation, in both the reproduction scenarios, is far from results obtained by Zahorik [258]. For this reason, our virtual resonating environment could be adopted in the setup of auditory displays where sounds in the far-field must be presented, without any particular requirement on the reproduction device.

In this chapter, a virtual listening environment capable of sonifying sources located at different distances has been presented, along with a versatile way to model it. Listening tests show that it actually conveys exaggerated range cues. Nevertheless the model can be easily re-parameterized to account for different psychophysical scales. In this perspective our model is prone to further optimization, as new reverberant spaces can be explored and straightforwardly validated through psychophysical experiments similar to those illustrated in this chapter.

Moreover, as stimuli for the two experiments, we used the impulse responses produced by our model, convolved with a short, anechoic sample of a cowbell. It would be interesting to repeat the

Figure 10.7: Loudspeaker listening: Individual distance evaluations together with individual linear regression lines. a: intercept. b: slope.

tests with another type of sound, in order to study how sound familiarity affects the data results.

Figure 10.8: Loudspeaker listening: Average distance evaluation together with linear regression line. a: intercept. b: slope.

Chapter 11

Complex gestural audio control: the case of scratching

Kjetil Falkenberg Hansen and Roberto Bresin
Kungl Tekniska Högskolan – Department of Speech, Music, and Hearing
Stockholm, Sweden
hansen@speech.kth.se, roberto.bresin@speech.kth.se

11.1 Introduction

To scratch means to drag a vinyl record forwards and backwards against the needle on an ordinary turntable along the grooves, not across, though it might sound like it. This way of producing sounds has during the last two decades made the turntable become a popular instrument for both solo and ensemble playing in different musical styles, still mostly in the hip-hop style where Disk Jockeys (DJs) first started to scratch. However, all musical genres seem to be able to adopt the turntables into their instrumental setups. Composers in traditions like rock, metal, pop, disco, jazz, experimental music, film music, contemporary music and numerous others have been experimenting with DJs the past years. Experimental DJs and most hip-hop DJs now frequently call themselves "turntablists", and the music style of scratching and extensive cut-and-paste mixing is called "turntablism". These terms, derived from the word turntable, are now generally accepted. It is also generally accepted that a turntablist is a musician and that the turntable is to be considered an instrument. The acoustics of scratching has been barely studied until now. On the other end the business market of DJs equipment is quite large. It is therefore interesting to study the phenomenon of turntablism from a scientific point of view.

In this chapter three experiments are presented. Aim of these experiments is to model scratching based on analysis of an experienced performer. For this purpose scratching as an expressive musical playing-style is looked at from different views. Experiment 1 investigates the musical fundamentals of scratching, and explains the most common playing-techniques in more detail.

Equipment	Description		
Turntable	Technics SL-1210 Mk2 with felt slipmat		
Cartridge	Stanton 500 (Experiment 1 only) and Shure M44-7		
DJ-mixer	Vestax PMC-06 Pro		
Faders	Vestax PMC-05 Pro		
Record	1210 Jazz - Book of Five Scratches. Book 2 [125]		
Potentiometer	Bourns 3856A-282-103A 10K (*Experiment 2 only*)		
DAT-recorders	Teac RD-200T Multichannel (*Exp. 2 only*)	Channel 1 (20 kHz)	Potentiometer
		Channel 2 (10 kHz)	Crossfader
		Channel 3 (10 kHz)	Sound
	Sony TCD-D10 (*Exp. 1 and 3*)	2 channels (44.1 kHz)	
Wave analysis software	Soundswell Signal Workstation [64]		
	Wavesurfer [221]		

Table 11.1: Equipment used for the experiments.

Experiment 2 investigates a real performance with aid of sensors on the equipment in order to understand what kinds of problems and parameter variation a model will need to deal with. Experiment 3 investigates what happens to the record that is being scratched, as it is apparent that some heavy damage will be done to the vinyl and this can therefore affect the sound quality.

11.1.1 Common set-up for experiments

All the experiments shared part of the equipment and methods, as explained in this section. Setups and methods specific to each experiment are explained in the respective sections.

11.1.2 Method

In all the experiments presented in the following, the same method was used. Small exceptions in the equipment are presented in Table 11.1[1].

Subject

Only one subject performed during experiments 1 and 2, examining expressive playing. He is Alexander Danielsson, *DJ 1210 Jazz*, a professional DJ from Sweden. He volunteered for the experiments. 1210 Jazz (as he will be called throughout the paper) has no formal musical training, but has for almost 15 years been considered to be among the best turntablists in Sweden and

[1] The audio recorded in Experiment 2 was only intended to illustrate what happened to the sound, and acoustical properties were not analyzed in this investigation.

Figure 11.1: Instrument set-up with mixer on the left and turntable on the right side.

Europe, a reputation he has defended in DJ-battles (as competitions for DJs are called) as well as on recordings, in concerts, and in radio and television appearances. He has made three records for DJ use, so-called *battle records*, one of which was used during the recording sessions.

For experiment 3, the first author performed the test.

Material

All the recordings used in these experiments were done at KTH during 2001. The equipment used for the experiments is summarized in Table 11.1.

Instrument line-up

Mixer and turntable were placed in a normal playing-fashion with the mixer to the left. The turntable was connected to stereo-in on the mixer. The right channel was output to a DAT recorder, while the left channel was output to a headphone mixer so the DJ could hear himself. See Figure 11.1.

Calibration

We needed to have a turntable with constant rotation speed for the experiments. The spikes (labelled 1-10 in Figure 11.2) in the waveform in the lower panel come from a cut made across the record, marking every 360° with a pop. Since the distance between each spike was equal, the turntable motor by assumption generates a constant speed. The upper panel in Figure 11.2 shows the readout from a 10-rounds potentiometer in degrees.

Figure 11.2: Calibration of turntable motor. The upper panel shows the record angle in degrees from the potentiometer. The lower panel shows the waveform from a prepared vinyl record.

11.1.3 Background - Playing the turntable

Introduction to scratch music

Turntablism includes both *scratching*, using one turntable, and *beatjuggling*, using two turntables. Scratching is a typical solo-playing style, comparable to that of electric guitar. The musician expresses himself in intricate rhythmical patterns and in tonal structures. Beatjuggling sometimes has a rhythmical backing function to hip-hop *rapping*, but is also often played in expressive solo-acts, typically with a groove on one turntable. A history of turntablism and overview of the practice of using turntables as instruments has been reported in a previous paper [114].

In the early eighties, scratching DJs and the entire hip-hop community got the attention of the record-buying generation, gradually spreading their sounds to the rest of the musical world. Since then, DJs have appeared on recordings by artists as diverse as John Zorn [259], Herbie Hancock [112], Mr Bungle [41], Portishead [196], David Byrne [45] and Tom Waits [248]. The first experiments with use of phonograph to create music started about 1915 and were carried on through the next three decades by Stephan Wolpe, Paul Hindemith, Ernest Toch, Percy Grainger, Edgar Varèse, Darius Milhaud and Laszlo Moholy-Nagy [56]. None of these composers were turntable artists, nor did they write compositions for turntable that still exist (although some compositions may be reproduced by memory). John Cage [47] and Pierre Schaeffer [213] composed what are considered to be the first pieces for turntables, but they are recognized mainly for their approach to sounds, not to the turntable. Pierre Schaeffer's concept of *musique concrete* is as much a result of tape-manipulation as vinyl-manipulation, and his practices would probably have been commenced even without the phonograph technology [107, 136].

Instrumentalists must learn to incorporate a variety of techniques and methods for tone manipulation in their playing, and DJs have established a fundamental ground with an almost compulsory assortment of techniques [115]. All new ways of playing and scratching are persistently explained and debated on, especially on Internet discussion forums on sites devoted to turntablism [215]. Some of the accepted techniques will be described and analyzed in Experiment 1. Numerous techniques more or less resemble, and often originate, from those described, but they will not be included in this study. Many ways to scratch do not fit into the general scheme, and are not widely used. They are, however, all explained on turntablism Internet sites as soon as they surface. The overview in Experiment 1 will not deal with the more unconventional scratches.

11.1.4 Equipment and instruments

Any turntable might be used for scratching, but the standard is a direct-driven machine with the platter mounted on the motor. Scratching also utilizes an audio mixer with a volume control that is mostly used for controlling the onsets and offsets of tones. Therefore, the turntable as a musical instrument includes the turntable with pick-up, slip-mat, mixer and a vinyl record. Relatively few manufacturers have succeeded to enter the market with highly qualitative equipment, considering the massive interest in DJing and the prospective of good sales. Turntablists seem to be skeptical to adapt to radical innovations, especially those that oversimplify and trivialize playing. One example is the production of a mixer that allows for some of the hardest techniques to be played with ease. This particular mixer, Vestax Samurai [241], never got much positive attention from the professionals, even though it could inspire to the development of new techniques. Why this is so and which consequences this inflict on the evolution of the scene would bring us to an extensive discussion. In brief, displacement technology could be considered to cheapen the musician's skill, and this explains the hesitation to switch to such a technology [104].

Low budget equipment fails to convince users. At a professional level there are several brands to choose from, that have comparable qualities. The standard line-up for many turntablists is a set made of two Technics SL-1200 Mk2 turntables with Shure M44-7 pick-ups, and a scratch mixer without unnecessary controllers, e.g. the Vestax PMC-06 Pro. In the following, "Technics" will refer to the Technics SL-1200 Mk2 or its black counterpart SL-1210 Mk2. Normally this equipment is organized with one turntable placed on either side of the mixer, and even rotated 90° with the tone-arms away from the body to avoid hitting the tone-arm with a botched arm movement.

The quartz direct-driven Technics (Figure 11.3) has advantages to the belt-driven turntables in terms of pick-up speed and motor strength (the starting torque of a Technics is 1.5 kg/cm). When pushing the start-button, a mere third of a revolution (0.7 s) is needed to reach full speed at $33\frac{1}{3}$ revolutions per minute (rpm), and stopping the record goes even faster [224].

With a felt slipmat lying between the platter and the record, moving the record in both directions with the motor on and the platter still going is fairly easy. In rougher moves and when maintaining a heavier touch, the platter might follow the record movement, but because of the pick-up speed, that is not a problem.

The *pick-up* mechanism is specially designed for scratching to prevent the *diamond stylus tip*

Figure 11.3: Technics SL-1200 Mk2 turntable (SL-1200 is silver gray and SL-1210 is black).

from skipping from one groove to another, and even the stylus tip is designed to make scratching easier. "Needle" is the popular term for both the stylus tip and the bar or tube that holds it, called the *cantilever*. Normal high fidelity styli are often elliptical at the end to follow the groove better, while scratch styli are spherical and made stronger as the force on the needle can get so high that both the reels and the needle are in hazard of being damaged. An M44-7 can handle three times the tracking force (the force holding the needle in the groove) of the hi-fi cartridge Shure V15 VxMR—up to 3.0 g (grams) on the M44-7 compared to 1.00 g on the V15 [220]. Output voltage from the scratch cartridges is 3-4 times higher (9.5 mV) than the hi-fi cartridge. The hi-fi cartridge has better frequency response, from 10 to 25 000 Hz, compared to M44-7's 20 to 17 000 Hz. On scratch cartridges the cantilever holding the stylus has high mass and stiffness, and the high frequency range is much affected and poorly reproduced. M44-7 has twice the cantilever diameter compared to the V15.

The *mixer*, see Figure 11.4, has other features than just amplifying, the most important ones are the volume controls and crossfader. DJ mixers were originally designed for switching smoothly between two turntable decks when doing beat mixing[2], and all mixers have for both stereo channels at least a volume fader and tone controls (bass, mid and treble ranges). The crossfader was developed during the late seventies to make beat mixing easier, and is positioned at the front end of the mixer. In Figure 11.4 the close-up shows crossfader and volume faders.

The crossfader glides seamlessly from letting in sound from channel 1 (left turntable), through mixing signals from channel 1 and channel 2, to channel 2 (right turntable) only. Whether a left-positioned crossfader shall let in sound from the left or right turntable is now adjustable with a special switch called *hamster switch*. Fading curves for crossfader and volume faders can be customized on most mixers. Several varieties in how the crossfader operates are fabricated, such as using optics or Rane's *non-contact magnetic faders* [198]. "Fader" might, however, be a misguiding term as the fading curve on the crossfader normally is adjusted to be very steep, making it act

[2] Eventually, better mixers allowed the traditional beat mixing of two different songs to evolve into the practise of making one section of a record last for as long as wanted by cross mixing between two identical records

Figure 11.4: Rane TTM56 scratch mixer with a close-up on crossfader and volume controls.

as an on-off switch.

Any vinyl record might be used for scratching, but during the last ten years sounds with certain qualities have dominated. Only fragments of original recordings are being manipulated, often shorter than one second. Isolated musical incidents like drumbeats, guitar chords, orchestra hits and especially sung, spoken or shouted words represent the majority of sounds used. To simplify playing, and for obvious economical reasons, a number of popular sounds are compiled on one record pressed for DJ use, commonly called a *battle record* (copyrighting is understandably an abiding but disconcerting question). The sound samples come without long gaps in one track. Longer sentences are often scratched one word or syllable at the time.

Sounds can be given different onsets and offsets depending on the technique used by the DJ. These variations on the sounds and the results of different techniques applied to the sounds will be described and analyzed in the following section presenting the Experiment 1.

11.2 Experiment 1 - The techniques of scratching

11.2.1 Introduction

The aim of this first experiment was to investigate and explain some of the different scratch techniques.

11.2.2 Method

Material

DJ 1210 Jazz was asked to perform some typical techniques, both independently and in a natural musical demonstration, and he also included more unusual examples. Further he was restricted to one frequently used sound, the breathy-sounding *Fab Five Freddie* “ahhh”-sample from “Change the beat”(a chorused voice sings “ahhh” with a falling *glissando* in the end [81]). The recordings lasted for about 30 minutes, out of which long sections (several seconds) of continuous playing of a single technique were extracted. About twenty different techniques were recorded in this session.

11.2.3 Techniques

Basic movements

Basic DJ hand movements are naturally distinguished in two separate entities; record movement and mixer (crossfader) movement. Most scratching techniques derive from fast but rather simple movements. Record control and mixer control depend strongly on one another, analogous to right- and left-hand movements in guitar playing. Both hands can operate on both devices, and most players switch hands effortlessly, both for playing purpose and for visual showing-off purpose. Even techniques where both hands go to the same device are performed. A general rule has been to have the strong hand on the vinyl, but with a more intricate use of crossfader, learners now tend to use their strong hand on the mixer instead. The volume controls (first and foremost the crossfader) are handled with fingertips and are often bounced swiftly between the thumb and the index finger. The record is pushed forwards and backwards in every imaginable manner. Travelling distance for the vinyl varies from less than 10° to more than 90° in each direction.

Onsets and offsets depend on the position of the sample when starting a scratch, the use of crossfader, and the speed of the movement. Three fundamentally different onsets are possible to achieve. First, a scratch movement can start before the sound sample, and the acceleration then completes before the sound cuts in. The crossfader will have no effect. In the second kind of onset, the scratch movement starts within the sound sample without use of the crossfader; the record will speed up from stand still and produce a very fast glissando from the lower limit of auditory range to desirable pitch, often above 1 kHz. A third onset category occurs when the crossfader cuts in sound from within the sample, creating an insignificant *crescendo*-effect, as if it was switched on. Any sound can be deprived of its original attack by cutting away the start with the crossfader.

Figure 11.5 shows a sector of a record with a sampled sound on it. The y-axis represents the position in the sample. There are eight different forms of a forward-backward motion, marked (a)-(h). All movement types are permutations of starting, turning and stopping either within or without the sampled sound. Movement types (a)-(b) and (e)-(f) start before the sound, and movements (c)-(d) and (g)-(h) start within the sound. Movements (a), (c), (e) and (g) have the change of direction outside the borders of the sound, while (b), (d), (f) and (h) change direction within the sound.

Figure 11.5: 8 ways of starting and ending a movement over a sound on the record.

Movements (a)-(d) end outside the sound border, while (e)-(h) end before the sound has finished.

Four of the possible movement types have been generated with the scratch model described in section 11.A. A simple forward and backward gesture is applied to a sound file, producing examples of types (a)-(b) and (g)-(h). Spectrograms of these examples are drawn in Figure 11.6, and demonstrates how the onsets and offsets will vary from each case. The sound used is a flute-like sound with the fundamental at 500 Hz. Spectrograms are cut at 9 kHz as the most interesting part is the shape of the fundamental and the added noise band that simulates groove wearing in the model. All four gestures represented in the spectrograms are identical.

Besides the starting, turning and stopping points, several other factors have influence on the output sound of a simple forward and backward movement: Direction, or whether to start with a push or a pull, is an effective variable. The speed changes the pitch, and can be fast or slow, steady or shifting. In addition to controlling sounds with the record-moving hand, the crossfader gives the DJ the option to produce alternative onsets and offsets. The kind of sound sample on which all variables are executed greatly affects the result.

Figure 11.7 shows the waveform and spectrogram of the sound being scratched. All the other figures presented in the following do not include the waveform plot, as the spectrogram contains all information about the interesting characteristics of the sound examples discussed here. Figure 11.7 is an illustration of a very simple scratch (this specific one will be explained as *baby-scratching*) similar to type (b) in Figure 11.5. The scratch is approximately 0.4 s long and can in traditional musical notation resemble two sixteenth notes at about 75 bpm (beats per minute). The original sound ("ahhh") has a noise band with a broad maximum, inducing the perception of some pitch. Arguably the first sixteenth has a broad maximum around 1200 Hz and the second sixteenth around 2400 Hz, or an octave higher, but it is hard to establish a definite tonal phrase because of the *glissando* effects that can be observed in the spectrogram in Figure 11.7. The first sixteenth starts

Figure 11.6: Spectrograms of 4 different types of scratch movement.

abruptly when the sound cuts in (this is evident on the amplitude level figure), while the second sixteenth has a smoother attack with an increase of both frequency and amplitude.

Ending of tones can equally be divided into three categories, with the quick slowing down to a halt perhaps being the most interesting one. During a scratch performance, a big portion of the onsets and offsets come from directional changes of the record movement within the boundaries of a sound sample. Figure 11.8 shows a combination of movement type (e) followed by type (c) (see Figure 11.5), where the turns from going forward to backward are made just beyond the ending of the sound. The 0.8 s long scratch is built up by four sixteenth notes at 75 bpm. The first sixteenth has an abrupt attack, while the second and fourth sixteenths have a more smooth attack. The third sixteenth has a faster attack than the second and the fourth, but the sound is still achieved in the same way. The explanation of these differences lies in the speed of the turn of record direction. In the example in Figure 11.7, the turn when going from forward to backward movement is quicker than the turn when going from backward to forward again—the initial move from the body is faster than the initial move towards the body. All the endings except the last one are results of slowing down the record to change the direction, producing a fast drop in frequency.

The pitch we perceive from the broad maximum of the noise band is determined by the original recording and by the speed by which it is played back. Normally $33\frac{1}{3}$ rpm is used for scratching, but

Figure 11.7: Waveform (upper panel) and spectrogram (lower panel) of simple forward and back motion, called baby-scratching

Figure 11.8: Spectrogram of two simple forward and backward movements.

also 45 rpm. These numbers can be adjusted by a certain percentage in both directions depending on the product: a Technics affords a detuning of the rotation speed by at most 8% [224]. 8% creates a span of almost a musical major sixth (from $30\frac{2}{3}$ to $48\frac{3}{5}$ rpm, a factor of 1.58). Perceived pitch is most influenced by the playback speed on the sample caused by the hand movements. There

Figure 11.9: Different hand positions taken from a performance by DJ 1210 Jazz.

are no musical restrictions (i.e. tonally or melodically) to which audible frequencies can be used, and no concerns about preserving the original pitch of the source recording. For a 500 Hz tone to reach 15 kHz, however, playback at 30 times the original speed, or 1000 rpm, is required, which is impossible for a human DJ to accomplish. Different source recordings cover the whole frequency range, and may even exceed the pick-up's range.

Hand motion

Each DJ has a personal approach to moving the record, even though the aim is a well-defined technique. There seems to be an agreement among performers on how it should sound, but not so much on how it is accomplished. Since the record has a large area for positioning hands and fingers, and the turntable can be rotated and angled as preferred, the movements can be organized with great variety (Figure 11.9).

The characteristics of the hand movements associated with different types of scratches will be examined in a future investigation.

Without crossfader

The most fundamental technique, also recognized as the first scratch, is done by pushing the record forward and backward, and without using the crossfader. When done in a steady rhythmical pattern of for example sixteenth-notes it is called *baby-scratch*. Movement types number (b) and

Figure 11.10: Spectrogram of tear-scratch.

(e) and the combination of (e) and (c) from Figure 11.5 are most frequent in baby-scratching. How fast the turntablist turns the record direction influences both attacks and decays. A long slowdown or start gives a noticeable glissando-like sound. In addition, the frequency-drop will make the listener experience volume decrease—this is thoroughly explained by Moore [179]. This evidence can also be extrapolated from equal loudness contours and *Fletcher-Munson* diagrams [75, 110].

Another fundamental technique is the *tear-scratch* that divides one stroke, usually the backstroke, in two separate strokes. The division is kind of a halt before returning the sample to the starting point. It is not necessary that the record stops entirely in the backstroke, but the fall in frequency and volume will give an impression of a new tone attack. Figure 11.10 shows how the simple division affects the sound.

Two techniques take advantage of a tremble-motion on the record. Tensing the muscles on one arm to perform a spasm-like movement is called a *scribble-scratch*. Dragging the record with one hand while holding one finger of the other hand lightly against the direction, letting it bounce on the record, make a stuttering sound called *hydroplane-scratch*.

Both hydroplane- and scribble-scratches produce continuous breaks in the sound. On a spectrogram of a hydroplane-scratch, Figure 11.11, it is possible from the tops and valleys around 1 kHz to trace how the finger bounces on the vinyl. The slowdowns at a frequent rate, here about 30 per second, produce a humming sound. The broad maximum makes jumps of about 1 kHz in these slowdowns.

With crossfader

The volume controls can cut a sound in or out at will, which was also the technique behind the experiments of Pierre Schaeffer conducted during the early fifties [191]. He discovered that

Figure 11.11: Spectrogram of hydroplane-scratch.

when removing the attack of a recorded bell sound, the tone characteristics could change to that of a church organ, for instance. Normally the turntablists let the crossfader abruptly cut the sound, in this way cutting the transition. The sound can easily be turned on and off several times per second, making the scratches sound very fast. This is probably one reason why scratching sounds so recognizable and inimitable. Some techniques are just baby-scratching with varying treatment of the crossfader. Others use longer strokes with quick crossfader cutting.

Forwards and *backwards*, *chops* and *stabs* all hide one of the two sounds in a baby-scratch, either the forward push or the backward pull. In *chirps* only two snippets of sounds are heard from a very fast baby-scratch. On every movement forward, the fader closes fast after the start, cutting the sound out, and going backwards only the last bit of the sound is included (cut in). At these points of the scratch, the vinyl speed is high and so the broad maximum of the noise band is high, 2 kHz in Figure 11.12. The drawn line, adapted from the baby-scratch spectrogram in Figure 11.7, shows the probable broad maximum curve of the forward and backward movement with 0.1 s silenced, hiding the change of record direction.

The most debated technique is the *flare*-scratch, with all its variations. Flaring means cutting out sound during the stroke, but discussions among the performers concern how many times and how regular these cuts should occur. In a relatively slow forward movement that starts with the sound on, the sound is quickly clicked off and back on by bouncing the crossfader between thumb and index finger. Various flares are given names based on the number of such clicks. A *2-click flare* is one stroke with two clicks, or sound-gaps, producing a total of three onsets. An *orbit* or *orbit flare* is the same type of scratch on both forward and backward strokes. In a *2-click orbit flare* there will be a total of six onsets; three on the forward stroke, one when the record changes direction and two on the backward stroke. The flaring-technique generates many other techniques.

Twiddle and *crab* further take advantage of the possibility to bounce the light crossfader between thumb and other fingers, making a rapid series of clicks. The superficial similarity with the *tremolo* in flamenco-guitar is evident. Figure 11.13 comes from a twiddle, where the index and middle fingers on the left hand thrust the crossfader on the thumb. The short gaps are easily audible even in a very short scratch.

Because the numerous clicks are done in one single stroke, the frequency of each attack or

Figure 11.12: Spectrogram of chirp-scratch. The overimposed grey curved-line shows the assumed record movement.

Figure 11.13: Spectrogram of twiddle scratch.

tone will be quite stable. Roughly speaking, *flutter-tongue* playing on brass instruments and flutes produces a similar sound [175].

The old technique called *transformer*-scratch is often compared to strobe lights, as the crossfader is pushed on and off fast and rhythmically during relatively long and slow strokes, or even at original playback speed. Now *transforming* is often performed with varying but simple rhythmical patterns on the crossfader, and thus controlling the record speed can be given more attention. Transforming generally attains greater tonal variety than the techniques where the main musical

Figure 11.14: Spectrogram of transformer-scratch.

purpose is producing short tones by rapid fader clicks, as, for instance, it happens in the twiddle-scratch. Figure 11.14 shows a spectrogram of a typical transformer-scratch section.

11.2.4 Discussion: Relevance in music

Techniques are seldom played individually for longer periods; conversely they are blended and intertwined for greater expression and musical substance. For instance, it might be difficult to distinguish one technique from the other during a two-click flare-scratch and a twiddle-scratch in succession. Indeed, the turntablists do not want to emphasize this. Rather, they often refer to the different playing styles and scratch solos in terms of *flow*, which, considered as a whole, seems to be more important than the components. Mastering one single scratch should be compared to mastering a scale (or, rather, being able to take advantage of the notes in the scale during a performance) in tonal improvisation. Without the sufficient skill, complicated patterns will not sound good at least to experienced and trained ears. Aspiring DJs, like all other musicians, have to devote hours to everyday training to get the right timing.

11.3 Experiment 2 - Analysis of a genuine performance

11.3.1 Introduction

In order to acquire knowledge about how scratching is performed and how it works and behaves musically, an analysis of several aspects of playing was necessary. Results from this analysis can be used as a starting point for implementing future scratch-models. By only looking at the individual technique taken from the musical context, it is easy to get an impression of scratching as a clean and straightforward matter to deal with. This is not always the case. Techniques are not often played individually, but rather shortened, abrupted, and mixed one with the other. Also many gestures are not necessarily classified as valid techniques, but as variations or combinations of existing ones.

Figure 11.15: Potentiometer set-up with top and lateral views of the turntable.

11.3.2 Method

In the DJ 1210 Jazz recording sessions eight performances were executed, all of which without a backing drum track. Since 1210 Jazz is an experienced performer, the lack of backing track was not considered a restrictive or unnatural condition even though scratching often is performed to a looped beat.

Equipment

A potentiometer was used to track the vinyl movement. The $3\frac{3}{4}$ rounds 10 kOhm potentiometer was mounted to the vinyl with the help of a stand, and a cylinder attached to the record center. The output was recorded by a multichannel DAT. The potentiometer was chosen based on how easily it turned. No effect could be noticed in the performance and friction on the vinyl when it was attached, and the DJ felt comfortable with the set-up. See Figure 11.15.

Modern mixers give the DJ the opportunity to change the fading curves of the crossfader. To get a reliable signal we decided to find the slider position from reading the output voltage, not the physical position. Two cables connected from the circuit board to the multichannel DAT recorder tracked the slider movement, but not automatically the sound level. The crossfader run is 45 mm, but the interesting part, from silence to full volume, spans only a distance of 2-3 millimeters, a few millimeters away from the (right) end of the slider run. Because the crossfader did not respond as the DJ wanted to, he glued a plastic credit card to the mixer, thus shortening the distance from the right end to where the crucial part (the so-called cut-in point) is located (see Figure 11.16). Positioned to the right, the crossfader completely muted all sound, and it let through all sound when moved a few millimeters (to the left).

Only the right channel of the stereo sound output signal was recorded to the multichannel DAT, but that was sufficient for evaluating the record movement output against the sound output. The original sound from the record had no significant stereo effects, and both right and left channel appear similar.

Figure 11.16: Crossfader adjustment. The white stapled-square marks the position occupied by a credit card used to shorten the slider range.

Calibrations

Both the crossfader and the potentiometer had to be calibrated.

To read the approximate sound output level from the position of the crossfader, every millimeter position was mapped to a dB level. A problem occurred as the slider had some backlash (free play in the mechanics). By using two different methods, both with step-by-step and continuous moving of the crossfader, the sound levels on a defined sound (from a tone generator) could be found and used as calibration for the output level. See Figure 11.17.

The potentiometer had a functional span of about 1220° or $3\frac{1}{2}$ rounds. Unfortunately it was not strictly linear, but we succeeded in making a correction to the output values so that the adjusted output showed the correct correspondence between angle and time. See Figure 11.18.

The dotted line in Figure 11.18 is the original reading from the potentiometer doing 3 rotations in 6 seconds, using the same method as for calibrating the turntable mentioned earlier. The dashed line is the correction-curve used to calibrate the readings. The solid line is the corrected original signal later applied to all recordings. The voltage was adjusted to rounds, expressed in degrees.

Material

The DJ was asked to play in a normal way, as he would do in an ordinary improvisation. He was not allowed to use other volume-controllers than the crossfader, but as the crossfader is by far the most used control during in a performance, and the other controllers are used to achieve the same sounding results, this does not affect the experiment. The performances from that session were by all means representative examples of improvised solo scratching with a clearly identifiable rhythmic structure; one of those examples is used here. 30 seconds of music were analyzed. All sounds were originated from the popular “ahhh” sound from “Change the beat” [81]. This sampled part is found on most battle-records, including the 1210 Jazz [125] record we used.

Figure 11.17: Crossfader calibration. The X axis shows the whole travelling distance of the slider in mm.

Figure 11.18: Calibration of potentiometer.

The analysis was done on the basis of three signals; the crossfader, the record movement and a waveform of the recorded sound, and for comparison even the audio track. Comparisons with previous recordings of the separate techniques provide valuable information on the importance of these techniques.

We decided to describe the music in terms of *beats* and *bars* in addition to looking at time. This description necessarily calls for interpretations, and especially at the end of the piece it is

Figure 11.19: Bar 7 transcribed to musical notation. The grey areas mark where the crossfader silences the signal. The upper panel is the low pass-filtered signal from the crossfader in volts, the middle panel is the audio signal, and the lower panel shows the rotation angle in degrees.

questionable if the performance is played strictly metrically or not. In this analysis, however, that is a minor concern. With our interpretation the piece consist of 12 bars in four-fourth time. The tempo and rhythm is fairly consistent throughout with an overall tempo of just under 100 beats per minute. Figure 11.19 shows an excerpt of the readings and illustrates how the structuring in beats and bars was done. The upper panel is the low pass-filtered signal from the crossfader in volts, the middle panel is the audio signal and the lower panel is the potentiometer signal in degrees. This excerpt is from bar 7.

11.3.3 Measurements outline

Vinyl movement

One of the things we wanted to measure was the movement of the vinyl record itself without considering the turntable platter or motor. The slipmat, placed between the platter and the record, reduces friction depending on the fabric and material. For these measurements the DJ used his preferred felt slipmat, which allowed to move the record quite effortlessly regardless of the platter and motor movement.

Crossfader movement

The second element we measured was the movement of the crossfader. To get a reliable signal we measured it directly on the circuit board.

Sound output

The third signal we recorded was the sound output from the manipulated record. In order to let the musician play in a realistic manner he was allowed to choose the sound to work with.

11.3.4 Analysis

In the following analysis some key elements will be considered; namely the work with the vinyl in terms of directional changes, angles and areas, speed and timing; the activity on the crossfader and the volume; the occurrences of predefined techniques; and finally the occurrences of different kinds of patterns. The three variables considered in the measurements are: (1) crossfader movements, (2) record movements, and (3) associated sound signal.

Sounding directional changes

Sound is obtained from scratching by moving the record forward and backward. This implies that the record will change direction continuously during playing. Directional changes can be grouped in three categories:

- changes which are silenced with the crossfader;
- silent changes, where the change happens outside a sound;
- changes where the sound is heard, here called *turns*.

Turns can be further categorized in terms of *significant* and *insignificant* turns, according to how well we can hear the directional change.

A significant turn will produce the attack of the next tone. An insignificant turn appears when only a little part of the sound from the returning record is heard, either intentionally or by imprecision, also producing a kind of attack (although less audible).

The record direction was changed 135 times, in average 4.5 times per second. Of such changes, 21.5% were heard: 18.5% of them were significant turns; 6% were insignificant. A technique like *scribble* would influence this result considerably, as it implies fast and small forward and backward movements (about 20 turns per second) with sound constantly on. This excerpt had two instances of short *scribble*-scratches, representing 36% of the significant turns. It seems that in a normal scratch-improvisation (at least for this subject), about 70-90% of the directional changes are silenced.

Further investigation is needed to explain why so many directional changes are silenced. More data from other DJs needs to be collected and analyzed. However, one possible reason could be that the characteristic and recognizable sound of a record changing direction is no longer a desirable sound among DJs wanting to express themselves without too much use of clichés. These characteristic sounds are typically associated with the early, simple techniques.

Angles and area

The length of a sample naturally limits the working area on the record for the musician, and moving the record forward and backward can be made difficult by the turntable's tone-arm. About a quarter of the platter area is taken up by the tone arm in the worst case. Big movements are difficult to perform fast with precision, resulting in a narrowing down, as the technical level evolves, to an average of about 90° (although not measured, recordings of DJs from mid-eighties seem to show generally longer and slower movements). We consider long movements those that exceed 100°. A little less than 50% were long movements.

The occurrence of equally long movements in both directions was quite low, about 30% of the pairing movements covered the same area. Only 25% of the forward-backward movements started and ended on the same spot.

Issues concerning rhythm and timing

An attempt to transcribe the piece to traditional notation will necessarily mean that some subjective decisions and interpretations have to be made. Nevertheless, some information can be seen more easily from a musical analysis. This transcription allows an analysis of timing in relation to the various scratching techniques, by looking at both the record and the crossfader speed and their relation to the corresponding waveform.

Speed

About half of all movements, both forwards and backwards, were done slower than the original tempo in this recording. The backward moves were often performed faster than the forwards moves (33% compared to 26%). Due to different factors, as inertia and muscle control, and the fact that scratching implies a rounded forward and backward stroke, it is hard to perform a movement with constant speed. Hence, most of the movements have unstable speeds and do not result in straight lines appearing at the potentiometer output.

Sound position

Even though a DJ has a great control over the record position, in this helped also by visual marks such as colored stickers, nevertheless a minor inaccuracy can affect the result greatly. 1210 Jazz had only one sound (and position) to focus on, so he did not make any serious mistakes resulting in unexpected attacks or silences. The sound sample was also quite simple to deal with. With continuous change of sound samples, or with sharper sounds such as drumbeats and words with two or more syllables, this issue becomes more problematic.

Crossfader

This analysis did not distinguish extensively between crossfader movements done with the hand or by bouncing with the fingers, but some evident cases can be pointed out. It is likely that the crossfader should be left open for performing a number of certain techniques, but the longest constant openings in this performance had durations which were shorter than half a second. The crossfader was turned or clicked on about 170 times in 30 seconds (more than 5 times per second). The total amount of sound and silence was approximately equal.

53.3% of the draws had one sound only, and 11.8% of the draws were silenced. Among the remaining draws, 24.4% had two sounds, 6.6% had three sounds and 3.7% of the draws had four separate sounds. Multiple sounds per draw were distributed quite evenly on backward and forward draws, except for the five draws carrying four tones; all were done on the backward draw.

Techniques

The aesthetics of today's musicians roots in a mutual understanding and practice of attentively explained techniques. However, the improvising does not necessarily turn out to be a series of well-performed techniques. So far, research on scratching has considered the performing techniques separately. A run-down on which techniques are used in this piece clearly shows the need for a new approach considering the combination of techniques and basic movements. All recognized techniques are here associated to the bar number they appear in. The duration of a bar is approximately 2.5 s, i.e. the DJ played with a tempo of about 96 bpm.

Forwards appear in the same position in almost every bar. There are 9 *forwards* in 12 bars; 7 land on the fourth beat (in bars 1, 2, 3, 4, 6, 10 and 12) and 2 *forwards* land on the first beat (in bars 6 and 9). All *forwards* on the fourth beat are followed by a pickup-beat to the next bar, except for the last *forward*.

Tear-like figures occurred from time to time when the sound was clicked off during the backward draw, but they do not sound as *tears* because the breakdown in the backward draw was silenced. 3 of these *tear*-likes are executed, in bars 6, 10 and 11. Normally, several *tears* are performed in series, and the sound is left on all the time. None of the *tears* here were clean in that sense, or perhaps even intended to be *tears*.

Chops normally involve a silenced return. Prior to 10 silences, a *chop* was performed. It happened in bars 3, 4, 5, 7, 8 and 11. A *chop* can be followed by another technique (but the whole forward move is used by the chop) as it happened during the experiment in bars 5, 7 and 11.

Stabs and *drags* are similar to *chops*, but performed with more force (faster). They both appeared in bar 8. Many movements (35%) had a swift crossfader use. There are two states of crossfader position during scratching: With the sound initially off, sound is temporarily let in; conversely, with the sound initially on, the sound is temporarily cut out. Main techniques of sound-off state are different *transform*-scratches, while *chirps*, *crabs* and especially *flares* are typical for sound-on state. Sound-on state should give more significant turns. Most of the significant (and insignificant) turns happened with variations on the *flare* scratch.

Some common techniques were not found in the recording of the performance under analysis, including *baby*, *hydroplane*, *chirp* and *tweak*. The reasons for this could be many; *baby* scratching will often seem old-fashioned while *tweaking* can only be performed with the motor turned off, so it is more demanding for the performer to incorporate it in a short phrase. The absence of *hydroplane* and *chirp* can be explained as artistic choice or coincidence, as they are widely used techniques.

Patterns

Some movements and series of movements are repeated frequently. Patterns are not considered to be valid techniques, and they are not necessarily so-called "combos" either. A combo is a combination of two or more techniques, performed subsequently or simultaneously.

Often a significant turn was followed by a silenced change and a new significant (or insignificant) turn in the experiment. This particular sequence was performed 6 times (in bars 1, 4, 6, 11, 12).

In the performance analyzed only 5 long (more than 100°) forward strokes were followed by another long forward stroke, and there were never more than 2 long strokes in a row. On the backward strokes, long strokes happened more frequently. 16 long strokes were followed by another long stroke; on three occasions 3 long strokes came in a row, and once 6 long strokes came in a row.

No forward stroke was silenced, while 16 backward strokes were silenced with the crossfader. As the *chop* technique involves a silenced return, this technique was often evident around the silences.

Two bars, bars 4 and 5, started almost identically, the major difference is that bar 4 had a *forward* on the fourth beat while bar 5 had a *chop* on the third offbeat.

Twin peaks

One returning pattern was a long forward stroke with a slightly shorter backward stroke followed by a new long forward stroke (shorter than the first) and the backward stroke returning to the starting point. This distinctive sequence looks in the record angle view like two mountain peaks standing next to each other, the left one being the highest, and as it returned 8 times in 30 seconds in this experiment, it was for convenience named *twin peaks*[3].

The *twin peaks* pattern was repeated 8 times with striking similarity. The first peak was the highest in all cases, ranging from 100° to 175° (132.5° in average) going up, and from 85° to 150° (120° in average) going down. The second peak ranges from 50° to 100° (77.5° in average) going up, and from 75° to 150° (128.75° in average) going down. All had about 10 crossfader attacks (from 7 to 11), and more on the second peak than the first. The second peak was always a variant of a *flare* scratch. The 8 *twin peaks*-patterns take up almost one third of the performance in time.

[3] After the TV-series by David Lynch called "Twin Peaks", with a picture of a mountain in the opening scene.

11.3.5 Discussion

The division and structuring of the recording into bars reveals that the techniques are used taking into account timing and rhythmical composition, such as fourth beats. For a better understanding of musical content in scratching, more recordings should be analyzed as only 12 bars and one subject do not suffice for formulating general conclusions.

11.4 Experiment 3 - Wearing of the vinyl

This experiment is a quite simple test to find the extent of wearing on a vinyl record used for scratching. During the first minute of scratching, a record groove will be drastically altered by the needle and start carrying a broad white noise signal.

11.4.1 Method

On 1210 Jazz's record [125] there is a set of sounds from Amiga and Commodore 64 games. One of these sounds, with a bright flute-like character, has a fundamental frequency at 600 Hz and a harmonic spectrum (harmonics at n·F0). This sound is repeated on the record in a short rhythmic pattern with small silent gaps.

On the very first playback of the record, high-end equipment was used to ensure highest possible quality. Second playback was done on the equipment used in the experiments, but with a brand new scratch needle. Third playback, which is used as reference in this paper, was made with the same needle after a few weeks of use (which should mean the needle is in close to perfect condition). After playing back the short segment of the record at normal speed, the record was dragged forward and backward over the sound for one minute, and then the segment was played back at normal speed again. The forward and backward movement was done none-expressive at a constant speed and with approximately 2 cm run on each side of the point on the record. This procedure was repeated over and over, so the final test covers 15 minutes of scratching (dragging the record forward and backward corresponds to normal scratch movements) and 15 playbacks of the same sound in different stages of wearing. All playbacks and scratching was recorded in stereo to DAT at 44 kHz sample rate.

Noise from the equipment

The noise coming from the equipment (especially the hum from mixer and turntable) is about 10 dB lower up to 4 kHz than the total noise emitted from a blank place on the record and the equipment. Figures 11.20 and 11.21 show the noise level from the record and the equipment with the first axis cut at 22 kHz and 6 kHz respectively.

Figure 11.20: Noise levels of recording (upper plot) and equipment (lower plot).

Figure 11.21: Noise levels of recording (upper plot) and equipment (lower plot) up to 6 kHz.

11.4.2 Results

The following figures show a selection of the spectrograms and line spectrum plots that were taken from every minute of the recording. Deterioration happens gradually, but only the most illustrating events are included here.

Spectrograms

Figure 11.22 show the original sound with surrounding silence before the scratching begins, and after 1 , 2 and 15 minutes of scratching. After one minute of dragging the record forward and backward, the signal clearly has deteriorated. Even the silent parts on each side of the sound signal start to carry a noise signal.

After 2 minutes of scratching, Figure 11.22, the whole surrounding silent part carries the noise

Figure 11.22: Spectrogram of the tone and surrounding silence after 0, 1, 2 and 15 minutes of scratching.

signal. The broad noise band seems to a have higher level of energy between 2 and 6 kHz. The upper harmonics (from the fourth harmonic upwards) that could still be seen reasonably clearly after one minute are from now on masked in the noise signal.

No big changes are detected at the following one-minute intervals until around the twelfth minute. After that, the tendency from the second minute spectrogram in Figure 11.22 of a stronger noise band between 2 and 6 kHz shifts toward being a narrower noise band (approximately 1 kHz) around 5 kHz.

After 15 minutes of scratching (Figure 11.22), the appearance of a narrower noise band is more evident. Below 2 kHz, however, not much happens to the original audio signal and the first two harmonics are strong.

Line spectrum plots

In the line spectrum plots in the following, only the pre- and post-scratching state (0 respectively 15 minutes) are considered. Line spectra taken before (Figure 11.23) and after (Figure 11.24) scratching show the same segment on the record. The harmonic peaks have approximately the same power, but a broad noise band is gaining strength.

The noise signal is more than 20 dB stronger after 15 minutes of scratching, which result in

Figure 11.23: Line spectrum of the tone after 0 minutes of scratching (0-20 kHz).

Figure 11.24: Line spectrum of the tone after 15 minutes of scratching (0-20 kHz).

a masking of the harmonics above 5 kHz. From the last spectrograms it seems that the wearing generates louder noise between 4 and 6 kHz. This noise band may be perceived as being part of the sound, especially with standard scratch sounds as “ahhh”.

The silent parts

The most interesting issue is the constant level of noise that will mask the audio signal, and the best place to look at the level and appearance of noise is in the silent parts surrounding the audio signal. The following line spectrum plots of silence before and after scratching illustrates to what extent the needle damages the vinyl record.

Silence is also affected when tearing down the grooves, in the sense that silence is replaced by a noise signal. Figure 11.25 shows the line spectra of the small silent gaps seen in Figure 11.22. Because the gap was short, less than 30 ms, a high bandwidth had to be chosen for the analysis. It seems that the noise is about 40-50 dB louder after 15 minutes for frequencies below 6 kHz, and

Figure 11.25: Line spectra of the silent part before (lower plot) and after (upper plot) scratching.

about 20-30 dB louder for frequencies above that.

11.4.3 Discussion

As noise seem to be generated in the groove already during the first minute of scratching, it seems unnecessary to consider a signal to be noiseless and perfect, at least if the level of realism strived for is that of vinyl being manipulated. A thing like a ‘clean’ signal will never be an issue in real scratching, which maybe also gives scratching its characteristic sounds. This same noise that appears in real performances can be implemented in a model for scratching, and maybe prove helpful in masking audible errors connected to resampling, which is often a problematic concern in today’s models.

11.5 Design issues for a control model for scratching

Considering the analysis from experiment 2, a scratch simulator must include a volume on/off function, as almost none of the scratches are performed with the volume constantly on. There is no need to be able to control bigger scratch areas than 360°, and 180° should be easily controlled. Probably a touch sensitive pad could be efficient for controlling the vinyl part. These are fairly inexpensive and have advantages compared other controllers. Finding some controller to match a real turntable will perhaps prove difficult and expensive due to the strong motor, heavy platter and the inertia.

To simulate the record playing, the sample to scratch should be looped. A sample prepared to be altered from a standstill state does not correspond to any real scratch situation, the closest would be a comparison with *tweak*-scratching, where the motor of the turntable is turned off, but then the platter spins easily with low friction. Many simulators today have the standstill approach. When the sample is running in a loop, a mouse may be used for dragging the “record” forward and backward. It will not feel much like scratching for real, however, as you have to press the mouse button on the right place on the screen and move the mouse simultaneously. Even if the ability to do this smoothly and efficiently can be trained, there are hopefully better ways. A touch sensitive pad is more suited to this task than both keyboards and mice. Since it registers touch, hold-down and release, it can be programmed to act as the vinyl would upon finger touch; a finger on the vinyl slows down the record easily to a halt without too much pressure, and the same can be achieved with touch sensitive pads.

From the analysis and data of the two first experiments a model of scratching was built using `pd`[4]. The readings of the potentiometer and the crossfader recorded in experiment 1 were used to control an audio file. By first using the output from the potentiometer to change the sample-rate of the audio file that was played back, and then using the output from the crossfader circuit board to change the playback volume level, we successfully resynthesized the few techniques we tested on. 3 techniques involved record movement only; *baby*, *tear* and *scribble*, while 2 techniques, *chirps* and *twiddle*, also involved crossfader movement. In the following, the measurements of these 5 techniques are analyzed in detail and some algorithms for their simulations are proposed.

11.5.1 Baby scratch model

An analysis of the data related to one baby scratch cycle (see Figure 11.26) shows that the DJ moves the record forward and backward to its starting position (0° in Figure 11.26a) in about 260 ms. The track used for the experiment was positioned at a distance of 9 cm from the center of the record. Thus it was possible to calculate the distance travelled by the record and the velocity of the record itself (Figure 11.26b). The velocity of this movement has the typical shape of target approaching tasks [74]. In the DJ pulling action, the velocity reaches its maximum when the record has travelled a little over half of the final distance, then velocity decreases to 0 value when the DJ starts to push the record forward. During the pushing action, the record increases in velocity (see the negative values of Figure 11.26b) in a shorter time than in the pulling phase. It thus reaches maximum velocity before having travelled through half the distance covered in the pushing action.

11.5.2 A general look at other scratching techniques

The Same observations and measurements can be done for the other scratching techniques taken in consideration in Experiment 2. In the following only the signals produced by the sensors are shown for each scratching type.

Both *chirps* and *twiddle* scratch models use *baby* scratch as the basic movement as do most scratches where the crossfader is the main feature. Still the record movement varies from the simpler *baby*.

[4] Pure Data, or `pd`, is a real-time computer music software package written by Miller Puckette (http://pure-data.org).

(a) (b)

Figure 11.26: Baby scratch: rotation angle of the record during one cycle (a); velocity of the record during one cycle (b).

11.5.3 Existing scratching hardware and software

Expensive equipment and hard-to-learn playing techniques are motivations for developers of turntable-imitating hardware and software. Several scratch simulators have emerged during the last ten years, but none have so far proven to be successful among the professionals. This is about to change, and one of the promising products today is the scratchable CD players that simulate record players by playing the CD via a buffer memory. This memory can be accessed from a controller. In the early stages of scratch CD players this controller was in the shape of a jog wheel, now it often is a heavy rubber platter that can freely be revolved. Recent models have a motor that rotates the rubber platter at an adjustable speed, making it resemble turntables even further. Buffer memory and scratch pad controllers are used for other media formats such as MP3 and MiniDisc as well.

The turntable itself is also used for controlling a buffer memory, either by attaching sensors or using the signal from the record. For the latter, standard pick-ups can be used, but the record signal must be specially coded. Attaching sensors to turntables has not yet been implemented in commercial products.

Software for performing scratching, often simulating turntables or making virtual turntables, is interesting above all for its low expenses and high versatility. The controllers are standard computer input devices or MIDI, but customable.

The game industry develop both simple and advanced applications and accompanying hardware that profit from the popularity of scratching. A common ground for all of this hardware and software, from the simplest on-line Flashwave game to the coded vinyl records, is the turntable.

(a) (b)

Figure 11.27: Tear scratch: displacement of the record during one cycle (a). Scribble scratch: displacement of the record during six cycles (b).

The inclusion of a crossfader-like volume manipulator should seem to be obvious, but so far it has not been dealt with satisfyingly.

11.5.4 Reflections

It is not obvious to see whether models of scratching will hold a comparison to vinyl technology. All simulators have in common their digital approach, which is quite natural, but there are benefits and catches with vinyl that are either overlooked or even sidestepped. One specific example of a vinyl-typical feature is the deterioration of the vinyl; a few minutes of dragging the needle continually over the same spot on the record has devastating consequences for the sound quality, and quoting experiences of DJs, the needle even responds differently to movement over that spot. CD players will not wear out grooves the same way a record player does, and this might take the edge off a sound the same way a producer in a recording studio can polish a rough performance to the nearly unbearable.

To simulate every aspect of the turntable, the vinyl, the needle *and* the more remote aspects like wearing, will probably turn out to be the only suitable option for making an acceptable replacement for today's instrument set-up. An approach built on physics-based modelling technique seems therefore appropriate and worth to experiment with in the future [206].

Arguably, the most characteristic quality in scratching is the big range of recognizable and universally agreed-upon playing techniques. Future research can reveal interesting issues regarding these. Also, technology aiming to replace the turntable should take into consideration the role

(a) (b)

Figure 11.28: Chirp scratch: displacement of the record during two cycles (a). Twiddle scratch: displacement of the record during one cycle (b).

and practises of scratch techniques. The techniques and characteristics of the hand movements associated with different types of scratches will be examined in future investigations.

11.A Appendix

The measurements conducted in the experiments reported in chapter 11 were used for the design of a model of scratching. The environment for this model is `pd`[5]. Sound models of friction sounds can be controlled by the scratch software, but it can also control recorded sounds in the same manner as turntables. The `pd` patch is open and customizable to be controlled by various types of input devices. We have tested the patch with both recorded sounds and physically modelled friction sounds, and we have controlled the model by use of computer mice, keyboards, MIDI devices, the Radio Baton, and various sensors connected to a Pico AD converter[6].

[5] Pure Data http://pure-data.org.

[6] Pico Technology. The ADC-11/10 multi channel data acquisition unit, http://www.picotech.com/data-acquisition.html.

11.A.1 Skipproof - a pd patch

Skipproof[7] has three main functions. It is an interface for manipulating the playback tempo of a sound-file by using a computer input device, and it is an interface for triggering models of single scratching techniques. With these two functionalities, Skipproof acts as both a virtual turntable and a scratch sampler/synthesizer. In addition, the output volume can be controlled manually or by the models of scratching as a significant part of some of the techniques. With this third functionality, Skipproof also simulates the scratch audio mixer.

Method

In addition to pd, Skipproof uses a GUI front-end program written for pd, called GrIPD[8]. pd processes all sounds, and GrIPD controls the different options made possible in pd.

Material

The sounds manipulated in Skipproof are 16 bit 44.1 kHz and 88.2 kHz mono wave-files, but other formats should easily be supported. Sounds, or 'samples' in analogy to DJ-terms, are meant to be 1.6 s long in order to imitate a real skip-proof record, yet there is no restriction to the length.

Apart from direct manual "scratch control" of a sound-file, it can be accessed via recorded scratch movements. These recordings originate from the measurements reported in the presented experiments.

Control concepts

Skipproof can be controlled by different sensors and hardware, and is easy to adjust to new input objects. Sensors, MIDI input and computer inputs (keyboard and mouse) are used both to trigger the scratch models and manipulate the virtual turntable and audio mixer.

Implementation

In the following, selected screenshots from the pd patch will be commented, explaining briefly how Skipproof is designed. pd allows the user to build complex patches with sub-patches and references to other patches. Skipproof is built up by many subpatches that send and receive control signals from and to one another.

Figure 11.29: Graphical interface for Pure Data, GrIPD with turntable and audio mixer controls.

GrIPD

Figure 11.29 shows the performance window in Skipproof. One main focus designing the graphical interface was to some extent copy a turntable and a mixer. There are also a number of other buttons and sliders not found on the standard hardware, enabling the DJ to change parameters of, amongst others, motor strength. The user will not have to open any other window than this interface just to play.

[7] The name Skipproof is taken from a feature found on DJ-tools records called a skip-proof section, where a sound (or set of sounds) are exactly one rotation long and repeated for a couple of minutes. If the needle should happen to jump during a performance, chances are quite good that it will land on the same spot on the sound, but in a different groove. The audience must be very alert to register this jump.

[8] GrIPD, or Graphical Interface for Pure Data, is written by Joseph A. Sarlo (http://crca.ucsd.edu/ jsarlo/gripd/).

Turntable and mixer

The large light grey rectangle in Figure 11.29 is the part that registers mouse action (when left mouse button is held down). The meter next to the grey area displays sound progression marked with each quarter round (90°, 180°, 270° and 360°). Around this 'vinyl record part' all the standard turntable buttons are collected; start/stop, two buttons for toggling 33 and 45 RPM, and a pitch adjustment slider. On a turntable there is also a power switch that lets the platter gradually stop rotating by its own low friction. When stopping the turntable with the stop-button it is the motor that forcefully breaks the rotation speed. The power-switch is sometimes used to produce a slow stop, but is omitted as a controller here.

Only two controllers are chosen from the audio mixer's many possibilities. The up-fader is a logarithmic master-volume slider. Normally the crossfader is far more utilized than the up-fader, but a slider is not an advantageous way to control volume when the mouse is occupied with record speed. Under the slider is a push-button which shuts out the sound (or turns on the sound) when activated. This button mixes the functions of the line/phono switch and the crossfader.

Other controllers

Top left in Figure 11.29, there is a main power-button "power up dsp" starting the audio computation and resetting or initializing some start values in the patch. Under this there are 5 buttons for selecting the playback sound (sample).

Under the heading "readymade scratches" there are several buttons for triggering the recorded techniques. Below, two sliders define the speed and the depth these techniques will have. The speed range is a bit broader than what is normally performed. The depth slider goes from long backward movements to long forward movements, also making exaggerated performances possible.

The button labelled "using volume switch" decides whether the recorded scratches should be performed with crossfader or an on-off switch.

The `pd` patches

The main window, Figure 11.30, opens the GUI and lists all the subpatches. The right side of the window receives the playback speed of the sound in rounds per second (rps). This information is sent to the progression meter in `GrIPD`.

The `pd trackselection` subpatch lets the user choose sounds and sample rates. All sounds will be read into the same table, "sample-table a", to reduce the CPU load.

"Metro" in `pd` is a metronome object counting at a defined speed. Here in `pd mousesens` the metro counts every 100 ms, and at every count registers changes in relative mouse position. Too slow or too fast movements (often caused by long time of no action or by inaccuracy in the mouse sensors) are filtered out. Both horizontal and vertical mouse activity is measured. The mouse speed value is sent to subpatch `pd loopspeed` for adjusting the playback sample rate.

The `pd loopspeed` is sent to the table-reader as seconds per round and can be altered by receiving the mouse speed, the pitch control value and on-and-off messages. When the motor is

Figure 11.30: Main patch for Skipproof.

turned off, the turntable will respond differently to movement caused by mouse activity. Some kind of inertia can be simulated, as in “Return to loop speed” in the top right corner of the window in Figure 11.32.

All the recorded scratches used for synthesizing single techniques are collected in `pd tables` for simplicity. Tables are read in `pd scratch-o-matics`. The empty “table11” is for techniques where the crossfader is not utilized, in this way all techniques can follow the same procedure in the subpatch described next.

Signals from crossfader and record movements are low-pass filtered at 30-50 Hz before implemented in Skipproof. Each of the techniques is picked out from a series of constantly performed single techniques, and so represent an idealized model. Apart from techniques where the main idea consists of many small movements on the record, as with chirps and scribble, only one stroke forward and backward is included for all scratches. The scratches vary in length.

In future versions, most of the tables will be substituted by algorithms for both record and crossfader movement.

The subpatch in Figure 11.34 reads the tables in `pd tables` using the same procedure, but

Figure 11.31: Mousesens: Calculating the mouse movement.

since the tables are of different sizes, some adjustments must be done to the computations. The record-movement table readings are sent to the main patch and replace the value from `pd loopspeed` in seconds per round. The crossfader-movement table readings are sent to `pd volcontrol` as both up-fader values and volume push-button values depending on which method is selected.

After a performed scratch, the turntable continues in the set RPM.

Figure 11.32: The loopspeed subpatch.

Noise and wearing

To simulate the wearing of the vinyl, as explained in experiment 3, a simple noise model was implemented, see Figure 11.35. Following several notions, it generates low-level white noise, narrow-band noise and vinyl defects as cracks and hiss. All the noise signals are recorded to a 1.6 s long table, so the vinyl defects always occur at the same spot on the record when it is looped.

11.A.2 Controllers

Here are some alternatives to standard MIDI and computer input controllers that we use. The model is controlled in 4 different areas. An analogue-digital converter from Pico sends the signals from the control objects to the computer. The voltage output is then read in `pd`, controlling the described parameters.

Figure 11.33: `pd` tables: The recordings of the techniques in tables.

Substituting the turntable

The Max Mathews' Radio Baton was used as gestural controller for Skipproof. The drumstick like batons were substituted by a newly developed radio sender that fits the fingertips. This new radio sender allows users' interaction based on hand gestures (see Figure 11.36).

Substituting the mixer

The crossfader on modern scratch mixers is becoming easier and easier to move; now some models have friction-free faders. Still it takes a lot of training to accurately jump the fader over the critical break-in point. To make it easier to accomplish fast clicks, a light sensor replaces the crossfader.

Figure 11.34: Scratch-o-matics: Reading the recorded techniques.

Substituting the record

Sounds in Skipproof are sampled sounds. The user can add her/his own choice of sounds to be scratched. Skipproof can also be applied to the control of the sound model, such as for friction.

Substituting the DJ

Max Mathews Radio-baton is divided into nine sectors, each sector hosting a pre-recorded scratch technique. Treated gloves (with wiring equivalent to the drumstick used with the Radio-baton) manipulate a 3D (xyz-plane) signal received by the antennas.

Figure 11.35: Noise generator.

Figure 11.36: The finger-based gestural controller for the Max Mathews' Radio Baton.

Chapter 12

Devices for manipulation and control of sounding objects: the *Vodhran* and the *InvisiBall*

Roberto Bresin, Kjetil Falkenberg Hansen and Sofia Dahl
Kungl Tekniska Högskolan – Department of Speech, Music, and Hearing
Stockholm, Sweden
roberto.bresin@speech.kth.se, hansen@speech.kth.se, sofia.dahl@speech.kth.se

Matthias Rath
Università di Verona – Department of Computer Science
Verona, Italy
rath@sci.univr.it

Mark Marshall, Breege Moynihan
University of Limerick – Interaction Design Centre
Limerick, Ireland
Mark.T.Marshall@ul.ie, Bridget.Moynihan@ul.ie

12.1 Introduction

In the case of physics-based sound models, the control of the synthesized sound is straightforward. As shown in previous chapters, these models offer direct access to sound source characteristics. In this chapter two applications that focus on direct control of sounding objects are described, the *Vodhran* and the *Invisiball*. These two applications have a common characteristic: they realize the control of a sound model by mean of physical gestures. This characteristic facilitated the design of the applications and the interface between the control model and the sound model.

The first application is the implementation of an augmented version of the *Bodhran*, called the *Vodhran*. The *Bodhran* is an Irish traditional percussive instrument. It is a relative simple instrument but characterized by good, even complex, expressive possibilities, and therefore it is a good starting point for the design of a control model. In the next section we illustrate control techniques based both on data obtained from measurements of gestures of drum player, and on the use of gesture controllers.

The second application presented in this chapter is based on a game, called the *Invisiball*. Users can move an invisible ball placed on an elastic surface by pressing the surface with a finger. The ball will roll towards the position occupied by the finger. The rolling of the ball is communicated to users by a rolling sound. The sound stops as soon as the ball has reached the finger position.

12.2 The virtual *Bodhran*: the *Vodhran*

In order to have a successful interface it is widely known that it is preferred to employ a metaphor that the end user of the artifact is familiar with. In the following pages we illustrate an application that aims to provide users with an expressive virtual musical instrument, based on the traditional *Bodhran*: the *Vodhran*. This is not designed entirely to simulate the *Bodhran*, but to provide an instrument which can be played in a similar way, and which creates a recognizable Bodhran-like sound. This instrument is to be an extension of the original, allowing for additional playing techniques and styles, which could not be accomplished with the real instrument.

In the *Bodhran*, sound is emitted in response to a stick (beater) beating the skin, generally controlled with the right hand. Sound is modulated/damped by pressing the inside of the *Bodhran* with the left hand. The beater is held loosely in the hand and is moved/controlled primarily by wrist action (rotation). The contact with the *Bodhran* is made with alternative ends of the beater in rapid succession. The damping factor depends on the left-hand pressure, and often a dynamic/colorful pitch range can be achieved by continuous damping control obtained by putting the left hand in counter direction to the beater. There is a variety of different applications of the damping factor employed by the musicians, e.g., fingers only, side of hand only and so on (see Figure 12.1).

In the following sections four control modalities are discussed. In all of them the same sound model implementing the *Bodhran* was used. This sound model is described in the next section.

12.2.1 Sound generation

The sound generation mechanism for the *Vodhran* is based on the (approximate and simplified) modal description of the drum and on the robust numerical solution of a nonlinear stick-membrane interaction model [11].

This approach aims at an integrated "sound object", oriented at the real drum, including different forms of a player's interference, rather than a perfectly realistic reproduction of isolated signals. The superordinate term of "modeling" points out this difference to sample-based sound production.

Figure 12.1: Traditional *Bodhran*.

Resonator-Membrane

The technique of modal synthesis [2] forms a well-suited basis for our efforts for several reasons (see Appendix 12.A for a description of the modal synthesis technique):

- real-time implementation requires a sound synthesis technique that delivers convincing results even under preferably low computational expenses - as opposed to, e.g., waveguide techniques;
- at the same time the possibility of dynamical interactions with the player, during changing position/velocity/direction of the stroke or variable damping gestures, must be provided (this demand addresses the basic drawbacks of sample playback techniques);
- the synthesis parameters should be comfortably estimable under physical and perceptional specifications as, e.g., tuning or intensity of damping. Modal parameters are particularly close to terms of auditory perception and can be estimated from guidelines to the aspired sound properties.

The practical procedure of defining and adjusting a modal synthesis unit modelled on the *Bodhran*

The sound of a *Bodhran*, struck at 6 equally spaced positions from the centre to the edge, was recorded. A relatively hard wooden stick with a rounded tip (resulting in a very small area of contact) was used and the strike was performed approximately perpendicular to the membrane with least possible external force applied by the player in the moment of contact (loosely held, resembling a pendulum motion).

The subsequent interaction between the stick and the membrane can be approximately seen as a one-dimensional impact, and the resulting curve of interaction force is close to a fed-in ideal impulse. Each excited membrane movement is therefore with good approximation handled as an impulse response of the resonator at the point of striking; its Fourier transform in turn approximates the frequency response. The modal parameters where finally extracted from these "frequency responses" according to the theory of the modal description [11]: Peaks in the magnitude response curves mark frequencies of resonant modes, decay factors are calculated from Short Time Fourier Transform (STFT) values at two different temporal points; the position dependent weighting factors [11] of the modes are given by the different peak levels (at each resonant frequency).

It is to be noted, that the described procedure shows of course many inaccuracies. In addition to the mentioned idealizations (spatially distributed strike interaction is not an impulse at one point, stick is not a point mass, not absolutely free from external forces) the signal recorded through a microphone does not match the membrane movement at one point, and peak values in the frequency response do not immediately display modal frequencies and weighting factors.

Our overall goal, though, is not a perfect imitation of the *Bodhran* sound, but a practically functioning expressive sound object inspired by the *Bodhran* in its main behaviour and sound characteristics. Under these premises even a first implementation sketch, under evaluation of only the lowest 16 frequency peaks, gave an agreeable result (b.t.w., also to the probably less computer euphoric ears of interviewed *Bodhran* players). An important final step consists of the tuning of synthesis parameters controlled by ear (in a final implementation together with sensory feedback), which remains the ultimate judging instance in our case.

The impact model

Our model of the impact interaction assumes a point mass with a certain inertia representing the stick, and an interaction force that is a non-linear function of the distance between stick and membrane [11].

The instantaneous cross relationship between the variables of the modal states (displacements and velocities), the state of the striker, and the force (expressed by an equation that contains the mode weights depending on the strike position), can—due to its "high degree" non-linearity—not be resolved analytically. Instead of inserting an artificial delay to solve this non-computability on the cost of new introduced errors (as commonly done), we are using an iterative procedure to numerically approximate the implicit variable dependency[1] [24].

[1] A (computationally) low cost version of the impact also exists; here the interaction force term is linear—excluding stickiness phenomena—and the resulting equation is solved explicitly, leading to a faster computation of variables.

Implementation.

The sound engine for the virtual *Bodhran* is implemented as a module for real-time sound processing software PureData (pd) [197]. Control messages (as incoming user "commands") are handled with the temporal precision of an "audio buffer" length; practical values for the pd internal buffer size are, e.g., 32 or 64 samples, providing a temporal accuracy of 32/44100 s ≈ 0.75 ms or $64/44100$ s ≈ 1.5 ms. This is perfectly acceptable even for this sensitive practical realization of a percussion instrument; latencies of sound cards/drivers and gesture control systems are much more problematic.

We end up with a flexible expressive algorithm that models interaction properties, i.e., the weight of the stick, hardness, elasticity and "stickiness" of the contact surfaces, as well as player's controls, i.e., strike-velocity and -position and damping, and efficiently runs in real-time in a standard environment.

12.2.2 Controlling the Vodhran

The sound model described in the previous section was controlled with four different modalities. These involve both software and hardware controllers (see Figure 12.2) , and are presented in the following.

Control by modelling real players

A pd patch implementing the model of professional percussionists was used for controlling the sound model of the *Bodhran*. This control model was obtained by the measurements described in chapter 7. The measured striking velocities of two players (subjects S1 and S2) were used as excitation velocities for the sound model. The pd patch, a "metronome rhythm generator", plays metronome beats with a striking velocity that is affected by dynamic level (*pp*, *mf*, or *ff*), chosen playing style (symphonic or drumset) and tempo. Any time an accent can be generated, and depending on the current tempo, dynamic level and playing style, the striking velocity exciting the model will be changed accordingly. The shift between the two playing styles can also be generated in real-time.

To enable a swift change between the different playing styles and dynamic levels these control parameters were connected to an external custom made switch board. The switches were connected to the PC through a device to take analogue signals through the parallel port, the pico [1].

The control for tempo was mapped to the Korg KAOSS pad, a MIDI controller with a touch pad. The x-axis on the touch pad controls the tempo between 100 and 250 beats per minute. The y-axis controls the fundamental frequency of the physical model. By using the touch sensitivity, from the touch pad it is possible to start the metronome by touching the pad and stopping it on release. In this way the onset and offset of beats becomes immediate and very flexible, resulting in less resemblance to a metronome. The rhythm generator was played in a live concert at Centro Candiani in Mestre-Venezia, Italy, on the 20th of June 2002.

Figure 12.2: The three hardware controllers used to implement the *Vodhran*. (A) The ddrum, (B) the Radio Baton, and (C) the Polhemus Fastrack.

The Radio Baton as control device

In the same concert, the metronome rhythm generator was used together with a more direct control device for playing, Max Mathew's Radio Baton [25].

The set-up used in the concert is shown in Figure 12.3. The set-up used the radio baton with the two radio transmitters at each end of a *Bodhran* stick, and the antenna was played with the stick as a "normal" *Bodhran* by the player, Sandra Joyce. The position of each end of the stick versus the antenna controlled the playing position of the sound model. During the concert the dampening and the fundamental frequency of the model was also controlled through devices such as the KAOSS pad and Doepfer pocket dial knobs. During the concert this was not done by the player herself, although this could be remedied by using foot switches that would leave the hands free for playing.

The ddrum as control device

In addition to the control devices described above, the ddrum [61] was used to play the sound model of the *Bodhran*. The ddrum is a commercial available electronic drumpad and the MIDI velocity out from the control unit was used to excite the sound model. For the pad used there was also MIDI poly-aftertouch, which was used to control the dampening of the model. The ddrum is a

Figure 12.3: Set-up used for the concert at Centro Candiani in Mestre-Venezia, Italy, on the 20th of June 2002. A player used a *Bodhran* stick, modified with cables to transmit radio signals, and played the antenna like a *Bodhran*. While the position across the antenna controlled the striking position of the model, the fundamental frequency and dampening was manipulated by another person through the Korg KAOSS pad and a set of knobs. The same person also played the "metronome rhythm generator" and then used the x-axis of the KAOSS pad to control the metronome tempo.

nice interface to play the model because of it's tactile feedback to the player and the lack of cables to the sticks used for playing.

Control by tracking gestures in real time

A fourth modality of control was based on tracking users body gestures. The process of tracking users movements requires a means of capturing gestures in real time, and extracting the relevant features of the gesture. This requires some form of input device which can take a gesture as input and extract the relevant characteristics of this gesture.

Traditionally, the use of gesture as input involves the use of computer vision techniques, such as the recording of gesture with a camera, and the tracking of the movement of a person over time. These systems have a number of limitations, for instance they can be too slow for real-time use, and do not cope well with tracking more than one user. Also, they might not be accurate enough for present requirements.

So, it was decided to use a Polhemus Fastrak[2] device in order to track the users movement. This system tracks the position of up to four small receivers as they move in the 3D space, with respect to a fixed-point electromagnetic transmitter. Each sensor returns full six degree-of-freedom measurements of position (X, Y, and Z Cartesian coordinates) and orientation (azimuth, elevation, and roll). The device connects to a computer through its RS-232 port, and operates at a speed of up to 115.2 Kbaud, with an accuracy of 0.03" (0.08 cm) RMS for the X, Y, or Z receiver position, and 0.15° RMS for receiver orientation, and a resolution of 0.0002 inches/inch of range (0.0005 cms/cm of range), and 0.025°.

The software system. The software architecture for the control of the Fastrak is made up of two layers. For the *Vodhran*, these layers have been implemented as `pd` externals.

At the bottom layer there is a driver, which allows the applications to communicate with the system, and to get the bare position and orientation information from each sensor.

This layer communicates with the Fastrak over the RS-232 connection using a binary protocol, in which all the information received from the Fastrak is encoded in binary format. This allows encoding of the complete position and orientation data for a receiver in just 14 bytes, and allows for receiving 1000 position readings per second. In conjunction with the Fastrak's update rate of 8 ms per receiver, this means that applications will receive every update from each sensor, and so record all available data on the gesture. This high update rate also means that applications should not have to wait for data at any stage, so that latency is minimized.

This bottom layer is constantly working, recording all current position and orientation data from each sensor, and making them available to the next layer up. This higher layer acts as a form of middleware for the system, taking the raw position data from the lower layer and extracting any necessary characteristics from this data. For instance, it calculates velocity, direction of movement, direction changes, etc. This allows applications to map specific characteristics of the gesture to parameters of the models directly.

Motion tracking. The Fastrak senses, for each activated receiver, two vectors:

- the position of a fixed point located inside the (receiver) sensor, referred to as receiver origin, is measured relative to the transmitter origin (which is analogously a fixed point inside the transmitter) in a system of (standard 3D) orthonormal coordinates (the transmitter system, which is in orientation rigidly connected to the transmitter);

[2] Polhemus Fastrak `http://www.polhemus.com/fastrak.htm`

- the orientation of the receiving sensor is expressed in term of the three angles (exactly their cosines) azimuth, elevation and roll. These values characterize three turning motions executed successively moving the transmitter system axes onto according receiver axes.

The coordinate change from receiver- to transmitter-system is accomplished by addition of the translation vector and multiplication with a 3D transformation matrix. While the translation vector is, of course, just the above position vector (or its negative), the transformation matrix has entries that are products of the direction cosines.

Calculations. For computational comfort the Fastrak allows immediate output of the transformation matrix entries themselves, so that the effort of their external calculation can be saved. What remains to be done is the matrix multiplication and (eventually) the translation (that is, an addition).

`pd` modules have been written, that execute matrix/vector multiplications and additions; they are combined in a `pd` patch to calculate the "absolute" position (i.e., relative to the transmitter) from receiver coordinates.

Geometry. In a first approach the receiver sensor has been rigidly fixed to a drumstick. The transmitter is used as an orientation system: It may be fixed to a frame indicating the imaginable or physical drum plane or might be possibly fixed to a part of the player's body that is not involved in the playing movement. Points on the drumstick can now be located within the transmitter system via the aforementioned `pd` patch. Of course the coordinates relative to the receiver, that is, dimensions of the stick and the fixing of the receiver, must be known to this end.

The tips of the stick are of course not of such small dimension that the point of contact with an assumed membrane (which would at first sight seem a necessary information) is the same for every stroke; it rather highly depends on the angle between stick and plane/direction of the stroke at contact time. A constant striking–point may not even approximately exist, since the head of a Bodhran stick may be around 2-3 cm in diameter. Tracking the position of a whole portion of the stick's surface and checking for distance from a striking-plane (or in a striking-direction) is of course computationally highly expensive in a real-time context.

It can though be noticed (see Figure 12.4) that for many Bodhran sticks (including the one we used) the tip is approximately spherical. As a consequence, a point inside the stick, at the center of an approximating sphere, is found at a nearly constant distance from a struck membrane for every stroke, independent from the striking-angle (and the actual point of contact). For all triggering strategies that we took into account, it suffices to track the position of such a point.

System design. In order to establish the requirements for the system, in terms of usability, methods of interaction and sound quality, a daylong workshop was held. Three expert *Bodhran* players, each one with their own distinct and different styles and techniques, took part. The players also had various amounts of experience in the use of technology in performance.

The players were asked to perform different traditional rhythms with the sensor attached to the beater (see Figure 12.5), and the results were recorded in order to further analyse the gestural patterns involved in *Bodhran* playing. Video data were also gathered for further analysis.

The results of the analysis of this data are being used to determine any common movements, gestures, or techniques performed by the players, so that the control parameters of the model may

Figure 12.4: *Bodhran* stick hitting a membrane at different angles. Note that the touching point varies, while the center point of the sphere forming the surface of the stick lies at the same distance from the plane for each stroke.

be extended in order to allow more interaction for the players.

By examining the data in this way, and continuing to extend the model, we ensure that the overall goal of the project is met, in that a virtual instrument is created, which can be played like a *Bodhran*, but is not limited to the physical characteristics of a *Bodhran*.

Analysis results

The analysis of the workshop has led to a number of additional features in the design of the system. The nuances of how a *Bodhran* is played are very complex, involving both the left and right hands, with the right hand beating the skin, while the left hand can be used to damp certain modes of the skin.

Damping. Left-handed damping was used by all players, and in some cases was used to produce very complex tone changes, even to provide a melody from the rhythmic beating of the skin.

This damping effect has a major role during the playing of the *Bodhran* and, as such, must be entirely incorporated into the system. Currently the sound model does contain a facility to damp the skin, but only at a single point at any given time. Also, the damping portion of the model would have to be coupled to the Fastrak hardware, and a Fastrak sensor attached to the left hand of a player, to allow them to damp the virtual Bodhran in a similar way to the real instrument.

Tactile feedback. During the course of the workshop, when the players were asked to use the beater from the virtual *Bodhran* to create a simple beat by striking a virtual plane, it was noticed that some players required more tactile feedback from the system than others.

While some players were able to hold the beat, using just the beater and listening to the gen-

Figure 12.5: Workshop session. Sandra Joyce playing the *Bodhran* with a sensor attached to the beater.

erated sound, one in particular found this difficult. The addition of a physical plane of reference, which matches the virtual one, was found to alleviate this problem. This opens to some further investigation, to determine whether or not a physical reference is required, and, if so, the form of this reference.

Frame of reference. Another point which was raised by this workshop was that of a frame of reference for the instrument. Currently, the system uses a fixed reference point, which does not move with the player. In order for any virtual instrument to be internalised there needs to responsive in a non-arbitrary way and the modification was made for an extension to expressivity and also to allow deliberate musical control on the part of the musician in terms of control sensitivity and control selection. Development based on the GRIP instrument—"gradually expand and personalize their gestural 'vocabulary' without losing acquired motor skills and therefore gradually add nuances to their performances without needing to adapt to the instrument" [182].

To meet this requirement, the system must allow the players to move naturally, as they would while playing the actual instrument. This would allow players to add to their movement range, without infringing on their standard movements.

To enable this, the frame of reference for the system needs to move with the player, so that should they turn or lean as they play (which most players seem to do), the system will continue to function normally, in its new frame of reference.

12.2.3 Conclusions

The control systems that were described in this chapter cover three different kinds of interaction with a sound model based on different levels of abstraction; (1) algorithmic control, based on measurements of real players gestures, (2) interaction by mean of percussive instruments, that allow players using traditional gestures with haptic feedback, (3) use of a gesture tracking system providing continuous controls to the sound model. In particular, the system based on the Fastrack is a novel gesture-based interface to a sound model, which is used as a virtual musical instrument. By basing the design of the system on the real instrument, and by involving players in the analysis and design of the system, it was our hope to create an instrument, which captures the intrinsic characteristics of the real instrument.

However, by not restricting the system to just the characteristics of the real instrument, and by not necessarily tying the instrument to any physical presence, an instrument which allows the players to expand their playing vocabulary can be created.

12.3 The *InvisiBall*: Multimodal perception of model-based rolling

"Rolling" sounds form a category that seems to be characteristic also from the auditory viewpoint. Everyday experience tells that the sound produced by a rolling object is often recognizable as such, and in general clearly distinct from sounds of slipping, sliding or scratching interactions, even of the same objects.

Prior experiments showed that listeners can discriminate differences in size and speed of wooden rolling balls on the basis of recorded sounds [120]. Results from perceptual experiments, performed in the framework of the Sounding Object Project (see chapter 4), demonstrated that listeners are able to perceive the size and weight of a ball under different conditions. These conditions were: (1) steel ball falling on a wooden plate, (2) wooden ball falling on a ceramic plate.

12.3.1 Controlling the sound of an invisible ball

The impact model implemented in the framework of the Sounding Object Project (see chapter 8) was controlled by the mechanic equations of a rolling ball. The sound model was controlled by simply feeding it with the X, Y, and Z coordinates of a target position in a 3D space.

In the demonstrative experimental setup presented in this paper, the space is given by the dimensions of a haptic interface that was constructed for controlling the Max Mathews' Radio Baton [25]. The default baton-like controller was substituted by a rubber thimble-controller. For this purpose a "thimble-sender" device was constructed (see Figure 12.6). The radio signal is sent by a fingertip. The interface is made of stretching material and it was placed over the receiving antenna. Finger position in the 3D space is detected in real-time and it is used by the algorithm controlling the rolling movement of a ball. By pushing the membrane with the "thimble-sender", users can

Figure 12.6: The “thimble-sender”; a controller for sending a radio signal from the fingertip to the receiving antenna.

make the ball rolling towards the finger by moving it from rest position in the 3D space (see Figure 12.7). The position of the rolling ball as a projection on the XY plane, i.e. as seen from above, is visualized on the computer screen. The position is represented as a colored disk assuming colors in the red-range at high speed (hot ball) and blue-range at low speed (cold ball).

This new interface allows users an interaction by using three different types of feedback:

- Acoustic - the sound model of the rolling ball.
- Haptic - control of the position of the ball by pressing the elastic membrane with a finger.
- Visual - graphical projection of position and speed of the ball.

12.3.2 Pilot experiment: multimodal perception test

The interface described in the previous section was used to test the quality of the sound model from a perceptual point of view. The main goal was to investigate the degree of reality of the rolling-ball sound when controlled with this interface. Since the interface allows three different feedback, acoustic, haptic and visual, three different experiments were conducted. The three experiments were run in parallel at the same time so that 3 subjects at a time could listen to exactly the same sound stimuli, in the way explained in the following.

Subjects and procedure There were twelve subjects, 6 female and 6 male. Their average age was 30. All subjects were researchers or students at the Speech Music Hearing Department of KTH, Stockholm.

Subjects listened to the sound stimuli individually over infrared headphones adjusted to a comfortable level. In this way three subjects at a time could take part in the experiment without seeing each other. The stimuli for each group of three subjects were organized as follows: (a) one subject could only listen to the sound stimuli, (b) another subject had both audio and visual stimuli, and

Figure 12.7: Haptic interface for the control of the Max Mathews' Radio Baton. The interface is placed over the receiving antenna. Finger position in the 3D space is detected in real-time.

(c) a third subject had both haptic and audio stimuli. Each subject was instructed to estimate the degree of realism for each sound stimulus. The responses were recorded on a scale on paper, from 0 to 10, with "unrealistic" and "realistic" as extremes. Stimuli were presented twice in a random order.

Experiment 1: haptic and acoustic feedback

Subjects. Subjects had an acoustic feedback through headphones and an haptic feedback from the finger-controlled interface presented in the previous section.

Stimuli. Nine sound model set-ups were used. They were obtained by combining 3 sizes of the ball with 3 damping values, thus producing 9 different set-ups. In this way the influence of both size and damping could be tested on the classification of the corresponding sounds stimuli as being realistic. These 9 set-ups were presented twice to the subjects. This correspond to a factorial design (3 sizes) x (3 damping) x (2 repetitions). By controlling the haptic interface with the "thimble controller" subjects produced 9 acoustic stimuli with 2 repetitions, for a total of 18 stimuli, of the duration of about 20 seconds each. The sound of each stimuli was that of a rolling ball.

Experiment 2: visual and acoustic feedback

Subjects. In this experiment subjects had an acoustic feedback through headphones and a visual feedback from a computer screen.

Figure 12.8: Effect of damping and size as resulted from the analysis of the responses in experiment 1.

Stimuli. The acoustic stimuli were those produced simultaneously at the same time by the subject controlling the haptic controller in Experiment 1. In addiction a visual stimuli was synchronized with the acoustic stimuli. The visual feedback was presented on a computer screen and it represented the real-time moving position of the rolling ball in the 2D space with speed/colour as third dimension.

Experiment 3: acoustic feedback only

Subjects. In this third experiment subjects had only an acoustic feedback through headphones.

Stimuli. The acoustic stimuli were the same as those in Experiment1 and Experiment 2. They were produced simultaneously by the subject controlling the haptic controller in Experiment 1.

12.3.3 Results and discussion

In the following a preliminary analysis of the results from the three experiments is presented.

A repeated measurements ANOVA was conducted on answers collected in Experiment 1. There was no significant effect of either size or damping parameters. The sound stimuli with low and medium damping produced by the model of a rolling ball of medium size were classified as being more realistic (see Figure 12.8). A high subjective variability emerged for stimuli obtained with a medium value of the damping factor. Values ranging from 0 to 8 were recorded.

Figure 12.9: Effect of damping and size as resulted from the analysis of the responses in experiment 2.

A repeated measurement ANOVA was conducted on answers collected in Experiment 2. There was no significant effect of either size or damping parameters. Nevertheless a closer observation of the results suggests that subjects tended to classify as more realistic the sound stimuli associated to low and medium damping and to large size of the rolling ball (see Figure 12.9). Also, subjects gave a higher preference rate to the most realistic stimuli as compared with best ratings given by subjects in Experiment 1. A high subjective variability emerged for stimuli obtained with a large value of the damping factor. Ratings ranging from 0 to 7 were recorded.

A repeated measurement ANOVA was conducted on answers collected in Experiment 3. There was no significant effect of neither size or damping parameters. A preliminary observation of the results suggests that subjects with only the acoustic feedback tended to classify as more realistic the sound stimuli associated to low and medium damping and large size of the rolling ball (see Figure 12.10). Subjects gave a higher preference rate to the most realistic stimuli as compared with best ratings given by subjects in Experiment 1. These results are comparable to those obtained in Experiment 2. A high subjective variability emerged for stimuli obtained with a small value of the damping factor and large size of the ball. Values ranging from 0 to 9 were recorded.

A repeated measurement ANOVA was conducted on all answers collected in the three experiments. The main results suggest a significant effect of the ball size parameter, F(2, 22)=6.6175, p=0.00562, and a significant effect of the damping factor of the sound model, F(2, 22)=5.5417, p=0.01124 (see Figures 12.11 and 12.12). There was no significant interaction between the size of the ball and the damping parameter.

Figure 12.10: Effect of damping and size as resulted from the analysis of the responses in experiment 3.

The average rate for each feedback modality across all stimuli is presented in Figure 12.13. The "acoustic & haptic" modality, Experiment 1, resulted to be the worst one, and the "acoustic & visual", Experiment 2, was classified as best, according to the ratings given by the subjects in the three experiments. There is no significant difference between average ratings of the "acoustic" modality and of the "acoustic & visual" modality.

Responses to stimuli can be averaged through all three experiments, as shown in Figure 12.14. It can be observed that, in all damping conditions, stimuli with small size are classified as less realistic. In general, stimuli with both medium size and medium damping were classified as more realistic.

12.3.4 Conclusions

In this section we proposed a new device for controlling physics-based sound models through direct manipulation of an haptic interface, the thimble controller based on the Radio Baton system. This interface allows a gestural control of a sounding object in a 3D space by direct manipulation. The "realistic" property of sounds produced by acting on the interface was analyzed in three experiments.

Main result was that sound stimuli corresponding to balls with large and medium size and low and medium damping were classified as more realistic.

As an overall result, sound stimuli were classified as more realistic by subjects using only

Figure 12.11: Effect of size obtained by averaging across all three experiments.

Figure 12.12: Effect of damping obtained by averaging across all three experiments.

“acoustic” feedback or “acoustic and visual” feedback. It seems that this is due to the difficulty in

Figure 12.13: Effect of modality on the classification of the sound stimuli in the “realistic” scale.

Figure 12.14: Average classification values for each sound stimulus used in the multimodal experiments.

controlling the haptic interface and the sound metaphor associated to it. Some of the subjects reported that it was difficult to imagine the ball rolling towards their finger. Nevertheless some of the subjects in Experiment 1 gave high rates to stimuli corresponding to balls of medium size. These results suggests that the haptic controller and/or the testing application can be better designed.

12.4 Acknowledgments

Our thanks to those workshop participants who helped in the analysis and design of the system - Tommy Hayes, Robert Hogg and Sandra Joyce.

12.A Appendix – Modal synthesis

Modal Synthesis is based on the description of the behavior of a resonating object in coordinates that are not displacements and velocities (or other physical state variables, e.g., flow/pressure) at spatial positions, but in terms of modes (these discrete *modal coordinates* correspond to the eigenfunctions of the differential operator describing the system. In the case of a finite linear system of lumped elements the modal coordinates can be calculated from the matrix connected to the finite system of differential equations describing this finite case) [2].

While lacking the immediate graphic meaning of the spatial state variables, the modal description of a vibrating object is of strong advantage in certain respect. First of all, the development of the system along each modal axis is independent of its state and development along the other modal coordinates (the differential equation of the system splits into a series of independent equations).

The free oscillation (that is, the evolution without external perturbation) of each mode can be analytically described, even in simple form: Each mode follows the evolution of an exponentially decaying sinusoid of a fixed frequency. The corresponding resonance behavior (i.e. the frequency response) is that of a lowpass filter with a peak around this modal (or resonant) frequency (the bandwidth of this peak is proportional to the inverse of the mode's decay time).

The spatial state variables of the system can, of course, be reconstructed from the modal states through a linear transformation: Concretely, the movement of a specific "pickup point"—giving the sound picked up in that point—is a weighted sum of the movements of the modes; conversely, an external input to the system at the pickup point (i.e., an external force) is distributed to the distinct modes.

Summing up, the full modal description of the system reduces to a series of mode frequencies with according decay factors. A series of weighting factors represents each interaction point (or practically speaking, each interaction point of possible interest). The transform function of the system with specific interaction—or pickup points—is finally a weighted sum of the above described resonance filters (just as the impulse response is the weighted sum of the described sinusoids). This finally shows the immediate (acoustic) perceptual significance of the parameters of the modal description that we gain in trade for the missing ostensive meaning of the modal coordinates themselves.

Based on the clear acoustic meaning of the modal formulation, simplifications in the implementation of the system can be accomplished such that they introduce the smallest audible effect on the sound; or the acoustic response may even, along with implementational complexity and computational cost, be simplified in an aspired direction. The modal approach in this way supports well the idea of audio cartoonification [96].

Chapter 13

Psychoacoustic validation and cataloguing of sonic objects: 2D browsing

Eoin Brazil and Mikael Fernström
University of Limerick – Interaction Design Centre
Limerick, Ireland
eoin.brazil@ul.ie, mikael.fernstrom@ul.ie

Laura Ottaviani
Università di Verona – Department of Computer Science
Verona, Italy
ottaviani@sci.univr.it

13.1 Introduction

In this chapter we describe a software tool developed in the Interaction Design Centre at the University of Limerick, which can be used for multiple purposes relevant to the Sounding Object Project. We also discuss two scenarios for psychoacoustic validation and for cataloguing of Sound Objects using the aforementioned software tool and we present the results of some experiments on these scenarios recently conducted at the IDC.

13.2 The Sonic Browser

From 1996 onwards, a software tool has been under development [28, 29] at the IDC at UL, for auditory and multimodal browsing. The initial aim was to explore if auditory display could be augmented by providing multiple and concurrent auditory streams in a highly interactive environment, supporting users to develop mental models of sonic information spaces. For musical content, the approach proved to be successful, and the concepts have been further developed and enhanced into the current version of the Sonic Browser. Originally, the main feature of the Sonic Browser was the ability interactively to choose and to listen to multiple audio streams of music concurrently. This concept has now been improved with a number of interactive visualisations and filtering mechanisms. The current version is also very promising as a client-server application that might be usable over local area networks and, perhaps, the Internet.

Currently, we see two main scenarios of use for the Sonic Browser in the context of the Sounding Object Project. Our first scenario is to use the tool for psychoacoustic validation of Sound Models. The second scenario is for sound designers to use the tool for managing catalogues of Sound Objects. In the following subsections we introduce briefly the two scenarios. In section 13.3 we describe the tool, while in section 13.4 and section 13.5 we present the results of the experiments conducted in the two contexts respectively. Finally, in section 13.6, we report our conclusions and future investigations.

13.2.1 Validation of Sound Objects

The development of Sound Objects in this project has been a progression of studies ranging from recordings of real sounds, listening tests, phenomenology, psychophysical scaling, to physical modelling of sound sources and actions.

In this chapter we outline an approach for validation of sound models, i.e. how Sound Objects compare to recordings of real sounds in terms of scaling, quality and realism, and we present the results of the recent experiments.

We have informally noted the difficulty of representing sound, both as recordings and as models, as the experience is always mediated either by microphones, recording/playback mechanisms and loudspeakers, or, by the models themselves. As the arguments about the relation between electronically mediated sounds and real sounds are outside the scope of the project, we henceforth disregard this somewhat philosophical discussion and look at the possibilities of empirical validation of Sound Models.

In the real world, we can quite easily determine the properties of auditory events, e.g. the size and distance to a bouncing object, its material properties, if it breaks (see chapter 4). Real sounds and people are situated in real acoustics, which helps us to pick up acoustic events. Sound Objects are, so far, not situated in an intrinsic environment and therefore they have to be seen as hyperreal representations of acoustic events, although in their final application they will be somehow encapsulated into objects and environments that will make them situated.

In section 13.4, we propose to empirically validate the properties and quality of Sound Objects using the Sonic Browser and by having the subjects sort recordings of real acoustic events and Sound Object events. In that section, we present the results of the experiments based on this approach and conducted recently by the authors.

13.2.2 Cataloguing Sound Objects

Cataloguing sounds is traditionally problematic. Many, or all, Foley artists know their collections of sounds and most often they appear to be indexed and sorted based on events and sound sources in text format. While this is the currently accepted practice for cataloguing sounds, it can be induced that it is quite inefficient. If you are working as a sound designer, reading about a sound, without hearing it or having a recollection of having heard the particular sound, is as productive as swimming a bicycle.

The same problem applies to other sonic workers such as electroacoustic composers, who work with collections of thousands of sounds and quite often think of sounds in an acousmatic sense, which makes verbal descriptions of sounds somewhat meaningless.

With the current version of the Sonic Browser, users can explore corpora of sounds by hearing. Visualisations are simultaneously available only to support interactive spatial organisation and indexing, based on the perceived auditory properties and qualities of the sounds.

In section 13.5, we introduce the results of the recent experiment in the cataloguing scenario.

13.3 The Sonic Browser - An overview of its functionality

The Sonic Browser current supports four visualisation mechanisms for the representation of audio data. These mechanisms are TreeMap, HyperTree, TouchGraph and SObGrid (Starfield display). These mechanisms are supplemented by the use of multistream audio and in certain cases an “aura” as well as three filter mechanisms.

A metaphor for a user controllable function that makes it visible to the user is the aura [71]. An aura, in this context, is a function that defines the user’s range of perception in a domain. The aura is the receiver of information in the domain, shown as a gray circle surrounding the cursor in Figure 13.1.

The filtering mechanisms are based on both the intrinsic and extrinsic object properties and include text filtering, dynamic sliders and colour filtering. The filters offer simple ‘And’/‘Or’ type filtering of the objects as well as the ability to load and save the current filter settings. It is important to state that all these mechanisms are dynamic and function in real-time. This work is based on Shneiderman’s mantra for design of direct manipulation and interactive visualisation interfaces “overview first, zoom and filter, then detail on demand” in mind [218].

Mappings that are used include acoustic/perceptual, onomatopoeia, source, sound type and event. Acoustic/perceptual mappings describe a sound’s physical attributes such as brightness, dullness or pitch. Onomatopoeia is where sounds are described by the way they sound, e.g. hammering could be “thunk-thunk” [249]. Source is the actual type of object associated with producing that sound or sounds. Sound type is used to break sound into various subcategories such as Speech, Music and Environmental. Audio events/actions are associated with what action or event has occurred, for example, a sound of a car braking would be “braking”.

Each object can be recognised by its visual and aural properties. The objects, which are under the aura of the cursors, are played simultaneously. Sounds are panned out in a stereo field

Figure 13.1: The Sonic Browser with SObGrid visualisation showing the aura in use.

controlled by the visual location of the tunes nearest the aura. The volume of the tunes playing concurrently is proportional to the visual distance between the objects and the cursor. The maximum volume will be heard when the particular object is centered under the aura. The auditory aspects are controlled by the aura or the cursor (if the aura is off or unavailable). The rest of this section will focus on particular aspects of the Sonic Browser's functionality.

13.3.1 The Sonic Browser Controls

The left side or control side of the Sonic Browser offers various buttons and mechanisms for dynamic control of the application as shown in Figures 13.2 and 13.3.

The slider is used to dynamically control the size of the aura. The buttons below this slider are the available visualisation buttons. These buttons activate the desired visualisation when pressed and display the particular visualisation in the right side/pane of the application. The next set of controls are the filter control mechanisms. They offer 'OR', 'AND', removal of the current filter settings, clearing the current filter settings as well as saving and loading of these settings. The current filter settings are displayed in the gray area to the right of these buttons. These control

Figure 13.2: The Sonic Browser with the HyperTree visualisation.

mechanisms are governed by the selections from the particular filtering mechanisms. The filtering mechanisms are shown directly below the controls and are dependent on the currently selected tabbed pane. These mechanisms are used to add particular filters and/or combinations of filters to the list of currently selected filters. The final area in Figure 13.3 represents the status area where information on the current user identification number, the connection state to the sound server and the connection state to the database are displayed.

13.3.2 The Sonic Browser visualisations

This section provides a brief overview of the available visualisations used in the Sonic Browser to illustrate these visual aspects and their purposes. Information visualisation techniques offer many components, which can be applied to a limited display surface used in accessing large volumes of data. There are two distinct categories: distortion-oriented and non distortion-oriented [151]. The Sonic Browser uses mechanisms from both categories. The techniques used are mainly from the branch of distortion-oriented techniques known as focus + context. Analyses of these techniques have been written by Noik [187] and by Leung [151], which we will not duplicate

Figure 13.3: The Sonic Browser Controls.

here. The various visualisations offer several viewpoints of the dataset. These allow users to gain a greater insight into the dataset by exploring through the various visualisations. Different domains required different visualisation approaches depending on the type of data which is being represented by offering multiple visualisations, a more domain-generic application is created.

Figure 13.4: The Sonic Browser with the TreeMap visualisation.

The TreeMap visualisation

The Treemap [127, 219] is a space-filling layout that is generated automatically, used primarily for overviews of document collections and their meta-data and is shown in Figure 13.4. Creating the Treemap involves dividing and subdividing the screen into rectangular areas of alternate horizontal and vertical divisions ("slice and dice") to represent subordinate nodes. The rectangles are colour coded to some object attribute to achieve the rectangular method used to display information objects. These rectangles can have associated text or labelling describing the information object it represents. A disadvantage of the Treemap is that, while being good in representing the attribute of the information structure portrayed through rectangle area (usually size), it is not so good at conveying the structure of the hierarchy.

The HyperTree visualisation

The HyperTree technique was pioneered at the Xerox Corporation by Lamping et al [147] and is based on Hyperbolic geometry which was one of the non-Euclidean geometries developed at the turn of the century [48]. It is a focus and context view that lays out the hierarchy dataset in a uniform way on a hyperbolic plane and then maps this plane onto a circular display region. It allocates the same amount of room for each of the nodes in a tree while still avoiding collisions because there is an exponential amount of room available in hyperbolic space. With a single still image, a projection from hyperbolic space looks similar to a Euclidean scene projected through a fisheye lens. The projection to hyperbolic space serves the purpose of a Degree of Interest function as required by Furnas [88]. The advantages of this technique are that any node of interest can be moved into the centre so its detail can be examined using simple dragging of the node. As this node becomes the focus of attention the entire tree is appropriately repositioned with this node as the root node. A node's context can be easily seen, as it is viewed from all directions of the tree with its parent, siblings and children shown in close proximity. The Sonic Browser HyperTree visualisation was already shown in Figure 13.2.

The TouchGraph Visualisation

The TouchGraph visualisation is a simple Zoomable User Interface (ZUI). ZUIs are based on alternative user interface paradigm using zooming and a single large information surface. The earlier foundations for ZUIs are found in work by Furnas and Bederson on multiscale interfaces called Space Scale Diagrams [89]. Detail can be shown without losing context, since the user can always rediscover context by zooming out. The TouchGraph visualisation uses geometric zooming where all the objects change only in their size so it simply provides a blown up vision of the dataset in combination with physics based forces. It also uses simple semantic zooming by selecting the number of related nodes and children of those nodes, which are displayed. The Sonic Browser TouchGraph visualisation is shown in Figure 13.5.

The SObGrid visualisation

The SObGrid visualisation is based upon the work by Williamson and Shneiderman [254] on dynamic queries. Dynamic queries allow for rapid, incremental and reversible changes to query parameters: often simply by dragging a slider, users can explore and gain feedback from a display in a few tenths of a second. The Starfield display is essentially a 2D scatter plot using structured result sets and zooming to reduce clutter. Ahlberg and Shneiderman [4] discuss this in the context of visual information seeking (VIS) which is a methodology especially created to support visual browsing tasks. The Starfield display provides an overview of the dataset with the information objects being represented by coloured dots or shapes. The properties of the information objects are used to map to the screen location, colour and shape for each of the objects and its related graphical representation. The number of visible screen objects is controlled through the use of dynamic queries; in the case of the Sonic Browser these are the various filtering mechanisms such as text filtering, dynamic sliders or colour filtering. The Sonic Browser SObGrid visualisation was already shown in Figure 13.1.

13.3.3 The Filtering Mechanisms

The Sonic Browser offers three different distinctive mechanisms for dynamic filtering of objects. The three mechanisms are the text filtering, dynamic sliders and colour filtering. These

Figure 13.5: The Sonic Browser with the TouchGraph visualisation.

mechanisms operate on information about the objects that have been stored in the underlying database. In the case of the text filter and the dynamic sliders this information is a set of text descriptions based on arbitrary classification of the object's properties. These descriptions of arbitrary classifications can be used in both classification and identification tasks.

Text Filtering

The text filtering mechanism uses user-specified properties of the objects to filter them. The possible shape properties of an object are shown below in Figure 13.6. The text filtering mechanism offers several categories of filtering on object properties from filename to music genre. It does not currently offer an automatic classification of the object to these categories but this idea has already been considered for a future interface mechanism as described by Brazil et al. [29].

Figure 13.6: The Sonic Browser’s Text Filtering Mechanisms.

Dynamic Sliders

The dynamic sliders mechanism is again based on the user-specified properties of the objects. The dynamic sliders are based on the Alphaslider [3]. Alphasliders are used for navigating a large number of ordered objects where the scale range corresponds to the range in which the items exist. The dynamic sliders can be described as a modification of the existing slider interface component to include both changing of the position within the scale and also of the scale itself.

Colour Filtering

The colour filtering mechanism is based on the current visual properties of the objects. This mechanism is more experimental. Its goal is to gauge the usefulness of filtering objects by their colour properties. It is a modified GUI component as shown in Figure 13.7 and is based on the normal colour chooser component. The preview pane is retained to ensure that the current colour is obvious. Further experimentation will validate this mechanisms advantages and disadvantages in the context of dynamic filtering of object.

Figure 13.7: The Sonic Browser's Colour Filtering Mechanism.

13.3.4 Control Mechanisms

The SObGrid component offers the ability to map a property of the object dataset to the X or the Y dimensions and as such requires a mechanism to perform this task. The Axis Editor is the component that performs this task. The current axis configuration allows only for a 2D display but the functionality for a 3D display has been included for a possible future release. The axis control mechanism is based on both arbitrary and non-arbitrary object properties to allow for a better exploration of the intrinsic properties of the objects. The aura within the Sonic Browser can be controlled by one of three mechanisms. These mechanisms are via the keyboard, via the Aura slider or via the Aura Size Dialog.

The Sonic Node Viewer Window allows for an object's properties to be accessed and modified dynamically. It can be opened by a control+right click on an object. The Sonic Node Viewer Window allows the user to browse all the properties of the objects stored in the current collection. The user can change any of the object's properties from the Node Viewer. It also includes a list of the other objects within the collection, which can be accessed by simply double clicking on their name within the list to update the Node Viewer and to show or edit the details of that particular object within the Node Viewer window. Accessing the related `pd` patches for any object within the Sonic Browser is easy. It can be opened by an alt+right click on an object with a related `pd` patch.

The Sonic Browser offers two methods of dragging object around the screen. The first method is a simple single object select; the second is a group object select which is currently only available in the SObGrid. Single objects can be moved around the screen by holding down the left mouse button while an object is directly under the centre of the aura.Group selecting object occurs where the mouse is clicked on the screen, but not on a object, and dragged to the appropriate point and released.

The Sonic Browser offers a simple method of tagging files of interest. An object can be tagged by simple choosing the ‘Tagged’ option in the Sonic Node Viewer Window or via the keyboard. The visual representation of the object is then updated with a halo around it, corresponding to the object current shape.

The Sonic Browser offers the ability to ‘Drag And Drop’ one or more sound files directly onto the application. These files will then be copied and placed within the current collection and will be fully accessible as with any other object within the collection.

13.4 Experiments in the validation scenario using the Sonic Browser

13.4.1 Brief summary of previous interesting approaches

A fundamental issue in conducting psychoacoustic experiments is the choice of the right type of test, in order to collect all the necessary data with the least amount of effort by the user.

Different approaches are suitable for different goals. We know the classical methods in psychoacoustics, such as the methods of limits, adjustment and constant stimuli. These methods lack sometimes certain features. Therefore, there is a wide research field trying to study and develop new approaches for psychoacoustic experiments.

This is why, for instance, the Signal Detection Theory was born, which focuses not on the stimulus thresholds estimation, but on the reasons underlying one particular subject decision. The SDT estimates judgement processes elements not considered by the classical psychophysical methods.

An interesting method is the Stimulus Sample Discrimination (SSD) method, proposed by Mellody and Wakefield [173], based on previous studies on the stimulus sampling procedure by Sorkin et al [227] and Lutfi [159, 160, 161]. It focuses on psychophysical discrimination experiments by using samples from a context distribution, which, according to the particular task, could be considered as additional information providers or as distraction components. The researchers report two main scenarios where this method could be applied: investigating on informational masking—that is the influence of the context on the listener’s decision—and studying the subject discrimination between two distributions. In their paper, they present an application of the SSD for evaluating the preservation of the singer identity in low-order synthesis.

Another important branch of experimental methods is classified as unidimensional and multidimensional scaling methods. They aim at estimating psychological scales which allow to compare the physical to the perceptual parameters, referring respectively to one and more dimensions.

An improvement of the classical multidimensional scaling technique is the one proposed by Scavone et al. [212], i.e. to use an interactive program, the Sonic Mapper, in order to allow the listeners to arrange, by drag-and-drop, the stimuli in a 2D space according to the similarities they find among them. The advantages of this approach are a decrease of the users fatigue, since it is possible to apply less comparisons than in the classical methods, and a consequent increase in attention and consistency of the listeners decisions. Moreover, as the users are able to compare all the stimuli interactively, they can appreciate the full set of stimuli more than in a pairwise comparison task.

13.4.2 The experiments

In the validation scenario we conducted a psychophysical experiment to compare real sounds to Sound Objects and to investigate the perceptual scaling of the physical parameters that control the sound models. The aim of the experiment is to begin to understand how the synthesized sounds produced by our models are scaled in comparison with the physical dimensions. We focused on two dimensions: perceived height of the object drop and perceived size of dropped objects. Our exploration was not limited only to the scaling task, but also encompassed the perceived realism of the event. Therefore, we divided the experiment in two phases, one concerned with the scaling task per se and the other one focused on the realism judgement. Moreover, as we wanted to compare the sound rendering of two different approaches in the sound modelling, the stimuli set included, besides recorded events, Sound Objects from both of these modelling approaches.

The experiment was preceded by a pilot probe, which used only one modelling approach. The pilot probe allowed for an initial observation of the type of results and it highlighted which sounds were most suitable to focus on in the main experiment.

The experiment used the Sonic Browser, that allows the users to listen to the stimuli and to drag-and-drop them according to their judgements within a specified bi-dimensional scale. The specified scale was a 2D plot with the perceived size of the dropped object on the X axis and the perceived height of the object drop on the Y axis.

The experimental data collection involved two techniques. First, data logging was collected by the application for the object positioning in the 2D space. Second, the user was asked to comment aloud on the thinking process, as it is established by the *Thinking Aloud Protocol.*

The Thinking Aloud Protocol is one of the most widely used methods in usability testing and it represents a way for the experimenter to have a “look” in the participants’ thought processes [22]. In this approach, the users are asked to talk during the test, expressing all their thoughts, movements and decisions, trying to think-aloud, without paying much attention to the coherency of the sentences, “as if alone in the room”.

We applied this protocol in our formative usability probe. Employing this protocol, we were able to collect not only the data concerning the stimuli positions in the 2D space of the Sonic Browser, but also the comments of the users during the experiment, which expressed the reasons, for instance, of a particular judgement or their appreciation of the Sound Objects realism. The tests were all recorded by a video-camera.

In the next subsections we will present both the pilot and the main experiment, introducing procedures and results of the test.

The pilot probe

Participants The pilot probe involved 4 volunteers, all students at the Computer Science Department of the University of Limerick. All of them referred to have neither hearing nor sight problems and all of them have a musical training (5, 4, 2 and 10 years respectively).

Stimuli The stimuli set included 9 recorded sounds and 9 Sound Objects, derived from the same modelling approach. All sounds, both recorded and synthesized, were not of a single collision, but of more than one bounce.

The recorded sounds were produced by 3 steel balls, weighting 6, 12 and 24 g, and falling on a wood board of 1500 x 500 x 20 mm from a height of 10, 20 and 40 cm, respectively, by positioning the microphone at 3 different distances: 20 - 40 - 80 cm, respectively. Recordings used a MKH20 Sennheiser microphone, and a sound card sampling at 44.1 kHz rate.

These stimuli were used in previous experiments conducted by the SOb project on the perception of impact sounds (see chapter 4). In this study, Grassi and Burro found the relationship between the physical quantities of weight, distance and height and the relative perceptual quantities. He argues that manipulating one of the physical parameters affects more than one of the perceptual quantities.

In the pilot probe, we decided to keep the height of the dropped balls constant (h=20 cm).

The synthesized sounds were all designed with the pd-modules modelling impact interactions of two modal resonators [9], simplified so as to return only one mode, and they used either glass or wood as the material property.

In Table 13.1 we report the list of the stimuli used in the pilot probe, with the relative short name and the sound type.

In Table 13.2 we report the values of the parameters used in the pd-modules for synthesizing the stimuli. For a reference on the main features of the models and the meaning of the parameters, we suggest to consult [9].

Procedure The probe was conducted in the isolation room of the Computer Science Department at UL. The stimuli were presented by stereo headphones to the user through the Sonic Browser. The experiment was conducted applying the Thinking-Aloud Protocol and the participants sessions were all recorded on video-tapes.

After the perception estimation task, the participants, during the second phase of the test, were asked to tag the sounds that they thought were unrealistic.

At the end of each session, a questionnaire was presented to the participants in order to gain an insight into their perceptions of the performed task.

The users estimated the data positions in the bi-dimensional scale without a comparison stimulus or a reference scale. Despite being pre-defined, i.e. being limited to the screen, the ranges

Short Name	Sound File	Sound Type
sound1	`d20-w12-h20.wav`	Real
sound2	`d20-w24-h20.wav`	Real
sound3	`d20-w6-h20.wav`	Real
sound4	`d40-w12-h20.wav`	Real
sound5	`d40-w24-h20.wav`	Real
sound6	`d40-w6-h20.wav`	Real
sound7	`d80-w12-h20.wav`	Real
sound8	`d80-w24-h20.wav`	Real
sound9	`d80-w6-h20.wav`	Real
sound10	`small-bouncing-glass-ball-1-pd.wav`	Sound Object - 1 Mode
sound11	`small-bouncing-glass-ball-2-pd.wav`	Sound Object - 1 Mode
sound12	`small-bouncing-glass-ball-3-pd.wav`	Sound Object - 1 Mode
sound13	`small-bouncing-glass-ball-4-pd.wav`	Sound Object - 1 Mode
sound14	`small-bouncing-glass-ball-5-pd.wav`	Sound Object - 1 Mode
sound15	`small-bouncing-wooden-ball-1-pd.wav`	Sound Object - 1 Mode
sound16	`small-bouncing-wooden-ball-2-pd.wav`	Sound Object - 1 Mode
sound17	`small-bouncing-wooden-ball-3-pd.wav`	Sound Object - 1 Mode
sound18	`small-bouncing-wooden-ball-4-pd.wav`	Sound Object - 1 Mode

Table 13.1: List of the stimuli used in the pilot probe.

of perceptual evaluations were relative to each user. The perceptual space boundaries were considered by all the users, as they reported at the end of the task, relative to their maximum value. In fact, we noticed an initial difficulty by the participants of referring to the screen space. On the contrary, they showed a preference of defining their own boundaries. In order to be able to compare the results of each participant, we decided to normalize the data coordinates, which identify the locations in the 2D space, between 0 and 1.

Results and Observations In Figure 13.8, we report the representation of the perceptual scaling and tagging information of all the users and all the stimuli in one graph. It is evident from this representation, that the participants estimate correctly the height from the real sounds, h=20 cm for all of them, since most of the real sounds, barring five outliers, are positioned in the central area of the evaluation space.

An interesting observation arises from Figure 13.9, which represents the individual perceptual scaling and tagging information sorted by users. Two participants in particular (users n. 2 and n. 3) made an obvious distinction between real and synthetic sounds.

In Figure 13.10, we represent the individual perceptual scaling and tagging information sorted by stimuli. As already mentioned, we can see that, while the height of the real sounds is perceived

Short Name	**Elasticity** k	**Damping** λ	**Gravity force**	**Strike velocity**	**Frequency** (Hz)	**Decay time** (s)
sound10	15000	46.4159	990	630.957	1758.52	0.043070
sound11	5540.1	8.57696	990	1318.26	1782.52	0.043070
sound12	15000	21.5443	950	1584.89	1388.82	0.090315
sound13	3161.6	21.5443	580	2290.87	1388.82	0.090315
sound14	15000	46.4159	450	630.957	1782.52	0.043070
sound15	15000	46.4159	450	630.957	1758.52	0.233307
sound16	15000	63.0957	940	912.011	1113.23	0.603386
sound17	11395	2.92864	860	301.995	1294.33	0.752992
sound18	1309.5	4.64159	970	436.516	1294.33	0.784488

Table 13.2: Values of the parameters for the synthesized sounds used in the pilot probe.

correctly, the size estimation varies to a degree between users. This could be influenced by either the distance and/or the conditions in which the real sounds were recorded.

Looking at the synthesized sounds, we noticed that for most of them, the participants agreed in the scaling task, at least for one of the two dimensions. Only for two stimuli (sound10, and sound16) the perceptual scaling was spread across the evaluation space.

It is interesting to look at the single stimuli, in Figure 13.11 which represents, through a box plot, the individual perceptual scaling of height and size respectively sorted by stimuli. For our purposes, we will focus on the synthesized sounds.

As it arises from these plots, two stimuli (sound10, and sound16) were hardly judged in a uniform way by all the users and the perceptual scaling is spread across the evaluation space, while the perceptual scaling of sound17 is slightly spread, but the latter was tagged as unrealistic by all the participants to the probe. Therefore, the data spread could be due to the lack of realism provided by the sound.

On the other hand, there is one sound, i.e. sound18, which is judged uniformly by all the users, and especially for the size dimension. We can see it more clearly in Figure 13.12.

Finally, the other five stimuli of the synthesized set were all judged uniformly in one dimension. In particular, sound11, sound12, and sound13, are perceived with the same height, while sound14 and sound15 are perceived with the same size. In Figure 13.13 and Figure 13.14 we report, respectively, the box plots of sound12, and sound15, examples of the uniform perceptual scaling in only one dimension.

We noticed that the participants did not agree in the results of the tagging task. Only one stimulus, sound17, was defined by all the users as unrealistic. On the other hand, only two, i.e. sound11 and sound13, were judged to be unrealistic by 3 participants. All the other stimuli received only two mentions. Moreover, it was observed that no participant identified any of the real sounds to be unrealistic. This is due to the presence of some reverberation (room acoustics) in the real sounds, that the synthesized stimuli lack.

Figure 13.8: Pilot probe: representation of the perceptual scaling and tagging information of all the users and all the stimuli.

Another interesting observation regards the type of approach taken by each participant to the task. Some of them preferred not to use the aura, but it was found to be useful for comparing the stimuli and checking the estimation of the whole group by those participants who used it.

At the end of each session as part of the participant debriefing, a questionnaire was presented to the participants in order to gain an insight into their perceptions of the performed tasks and about the Sonic Browser. A seven point Likert scale questionnaire with six sets of semantic differentials was filled out by the participants who were asked to express their responses to the interfaces and to the tasks, from 0 to 6, where 0 is "poor" and 6 is "excellent". In Figure 13.15, the results of the questionnaire with cumulative participant responses displayed per question can be seen with 0 representing a negative result to the question and 6 representing a positive result to the question.

Question 1 deals with the perceived difficulty in performing the task by the participant and the results show that it was found to be a non trivial task. Question 2 deals with the ease of use of the Sonic Browser for these tasks which was found to be above average ease of use. Questions 3 and 4 deal with the quality and realism of the sounds. The results from these questions show that

Figure 13.9: Pilot probe: representation of the individual perceptual scaling and tagging information sorted by users.

sounds were found to be both of high quality and realistic by the participants. Question 5 concerns a technical issue which arouse from the piloting phase of this experiment. There is a slight delay of up to 0.3 s when playing an audio file with the Sonic Browser. The result of this question was an acceptable but noticeable delay when playing sounds within the Sonic Browser.

The rich verbal protocol returned several interesting results during the experiment. The lack of room acoustics or background recording noise was commented by one participant who stated that some of the sounds did not have any "room effect". The most of the participants found that the "speed of bouncing was directly related to the realism of the sounds". The participants were found to use one of three strategies for scaling the sounds. These strategies were to "rough order the objects to the height scale first" or to "sort according to size initially" or to "sort them into real or synthesized sounds".

The aura was only found useful by half of the participants.

Figure 13.10: Pilot probe: representation of the perceptual scaling and tagging information sorted by stimuli.

The main experiment

Participants The participants to the main experiment were 5 volunteers, all students or workers at UL. All the participants referred to have musical training in average of 8 years, with a minimum of 6 and a maximum of 10 years. Two participants require glasses for reading, but no participant reported to have hearing problems.

Stimuli The stimuli set included 6 real sounds and 12 synthetic, 6 of which were designed with the `pd`-modules modelling impact interactions of two modal resonators [9], simplified returning only one mode, while the other 6 with the `pd`-modules modelling impact interactions of two modal resonators as well as the dropping event.

The real sounds were the same used in the pilot probe, but in this case we kept constant the distance (d = 80 cm), while changing the height.

Figure 13.11: Pilot probe: representation, by a box plot, of the perceptual scaling of the height and size sorted by stimuli.

For the synthetic sounds, that belong to two groups, we preferred to keep constant the material, since we noticed during the pilot probe some difficulties by the users to evaluate and compare the dimensions of events involving different materials. We decided on wood as the material, even if it is not clear if the wood is the material of the impactor or of the surface. In fact, even if the real sounds come from steel balls, they were referred to by the participants as wooden balls. This perception arouse from the bigger influence of the surface material in certain cases.

In Table 13.3, we report the list of the stimuli used in the experiment, with the relative short name and the sound type. As in the pilot probe, all sounds were not of a single collision, but of more than one bounce.

In Table 13.4 and Table 13.5 we report the values of the parameters used in the pd-patches of the two different parametrizations for synthesizing the stimuli sets. In Table 13.5, we report only the values that we changed for each sound. The other ones were kept constant at the following values: elasticity = 1e+007, alpha = 1.02882, lambda = 1e-006, strike velocity = -1.44544,

Figure 13.12: Pilot probe: representation, by a box plot, of the perceptual scaling of the height and size for stimulus n.18. Uniform perceptual scaling in both the dimensions.

minimum_& = 3.16228, maximum (regular) interval = 1000, multiplication factor = 0.88, interval deviation = 1, value deviation = 1. For a reference on the main features of the models and the meaning of the parameters, we suggest to consult [9].

Procedure The experiment was conducted in the same isolation room as the pilot probe. The stimuli were presented by stereo headphones to the users through the Sonic Browser. As in the pilot experiment, the Thinking-Aloud Protocol was applied and all the users performances were video-taped.

After the perception estimation task, the participants were asked to tag the sounds that they thought were unrealistic, and at the end, the participants were asked to fill out a questionnaire.

As in the pilot probe, the ranges of perceptual evaluations were relative to each user. We

Figure 13.13: Pilot probe: representation, by a box plot, of the perceptual scaling of the height and size for stimulus n.12. Uniform perceptual scaling in one dimension: the height.

decided to normalize the data coordinates between 0 and 1, for comparing the results of each participant.

Results and Observations In Figure 13.16, we report the representation of the individual perceptual scaling and tagging information sorted by users. As for the pilot probe and also in this case, we can see the classification by sound groups. Moreover, we notice that two of the participants (user n.1 and user n.2) only performed minor judgements on size of the real sounds. They referred, in fact, that they perceived other parameters changing, such as distance and material. This complex influence of the three parameters has been already discussed by Grassi and Burro in chapter 4.

In Figure 13.17, we represent the perceptual scaling and tagging information sorted by stimuli. The users take two different approaches to the two groups of synthetic sounds. In fact, with

Figure 13.14: Pilot probe: representation, by a box plot, of the perceptual scaling of the height and size for stimulus n.15. Uniform perceptual scaling in one dimension: the size.

the sounds synthesized by the first model, it seems that the users are much able to estimate the two parameters, since four of the six sounds (sound7, sound10, sound11 and sound14) are estimated coherently by most of the participants. On the contrary, the sounds designed with the second model are not clearly estimated and provide a spread of answers across the participants. A possible explanation of this spread of estimations may be due to the presence of a "buzz" tail, that conveys an unnatural perception of the event, and probably related to a low damping value in the parameter design, which causes the bouncing to persist up to "super-small" time intervals between two successive bounces.

It is interesting to observe the single stimuli, as we did for the pilot probe, looking at Figure 13.18 which represents, through a box plot, the individual perceptual scaling of height and size respectively for each stimulus. As with the previous experiment we again have focused on the

Figure 13.15: Results of the questionnaire for the pilot probe.

Figure 13.16: Representation of the individual perceptual scaling and tagging information sorted by users.

Short Name	Sound File	Sound Type
sound1	`d80-w12-h10.wav`	Real
sound2	`d80-w12-h20.wav`	Real
sound3	`d80-w12-h40.wav`	Real
sound4	`d80-w24-h10.wav`	Real
sound5	`d80-w24-h20.wav`	Real
sound6	`d80-w24-h40.wav`	Real
sound7	`small-bouncing-wooden-ball-1-pd.wav`	Sound Object - 1 Mode
sound8	`small-bouncing-wooden-ball-2-pd.wav`	Sound Object - 1 Mode
sound9	`small-bouncing-wooden-ball-3-pd.wav`	Sound Object - 1 Mode
sound10	`small-bouncing-wooden-ball-4-pd.wav`	Sound Object - 1 Mode
sound11	`small-bouncing-wooden-ball-5-pd.wav`	Sound Object - 1 Mode
sound12	`small-bouncing-wooden-ball-6-pd.wav`	Sound Object - 1 Mode
sound13	`w12-h10-pd.wav`	Sound Object - 2 Modes
sound14	`w12-h20-pd.wav`	Sound Object - 2 Modes
sound15	`w12-h40-pd.wav`	Sound Object - 2 Modes
sound16	`w24-h10-pd.wav`	Sound Object - 2 Modes
sound17	`w24-h20-pd.wav`	Sound Object - 2 Modes
sound18	`w24-h40-pd.wav`	Sound Object - 2 Modes

Table 13.3: List of the stimuli used in the experiment.

synthesized sounds.

We can observe that there is more uniformity in perceptual scaling in two dimensions, than in the pilot experiment. For instance, sound7 and sound11 have a strong uniformity in both the dimensions, despite an outlier for sound7 concerning the perception of its height. In Figure 13.19 and Figure 13.20, we report their individual box plots.

Considering other two stimuli, i.e. sound10, and sound14, (Figure 13.21 and Figure 13.22), we can see that there is a slightly uniformity in either dimension.

Four of the stimuli were perceived uniformly in one dimension, despite the presence of one outlier in most of the cases. Specifically, sound9 and sound15 were judged uniformly in the size dimension, and sound12 and sound13 were judged uniformly in the height dimension.

There were three stimuli with dispersed perceptions of their estimations (sound8, sound16 and sound18) and a highly dispersed perception of one sound in particular (sound17), whose individual box plot is represented in Figure 13.23.

Contrary to the results of the tagging task in the pilot probe, no stimuli in this experiment were tagged by all the participants. The maximum users consensus, regarding unrealistic stimuli, was achieved by 3 users. The real sounds were perceived again as realistic, duplicating the results of our pilot study.

At the end of each session as part of the participant debriefing, a questionnaire was presented to

Short Name	**Elasticity** k	**Damping** λ	**Gravity force**	**Strike velocity**	**Frequency** (Hz)	**Decay time** (s)
sound7	15000	46.4159	450	630.957	1758.52	0.233307
sound8	15000	63.0957	940	912.011	1113.23	0.603386
sound9	11395	2.92864	860	301.995	1294.33	0.752992
sound10	1309.5	4.64159	970	436.516	1294.33	0.784488
sound11	1309.5	8.57696	990	1318.26	1254.95	0.784488
sound12	3162.28	25.1189	900	524.807	1322.83	0.233307

Table 13.4: Values of the parameters for the synthesized sounds with the first parametrization.

Short Name	**Hammer mass**	**Initial interval** (ms)	**Acceleration/ Deceleration**	**Initial value**
sound13	0.0215443	228.530	0.76	0.56
sound14	0.0215443	306.516	0.74	0.75
sound15	0.0398107	207.223	0.72	0.57
sound16	0.0398107	207.223	0.72	0.57
sound17	0.0398107	277.939	0.70	0.75
sound18	0.0398107	372.786	0.70	1.00

Table 13.5: Values of the parameters for the synthesized sounds with the second parametrization.

the participants in order to gain an insight into their perceptions of the performed tasks and into the Sonic Browser. A seven point Likert scale questionnaire with six sets of semantic differentials was filled out by the participants who were asked to express their responses to the interfaces and to the tasks, from 0 to 6, where 0 is "poor" and 6 is "excellent". In Figure 13.24 and in Figure 13.25, the results of the questionnaire with cumulative participant responses displayed per question can be seen with 0 representing a negative result to the question, and 6 representing a positive result to the question. In Figure 13.25, we report three additional questions that were added to the questionnaire after the pilot probe and that asked about the learnability, interpretation of the application as it applied to the task and the difficulty in replaying the last sound.

The results of the first five questions show that it was found to be a non trivial task and that the Sonic Browser was found have only an average ease of use. Moreover, the sounds were found to be of a high quality but they did not seem particularly realistic to the participants. This can be attributed to the inclusion of two different types of sound objects containing either one or two modes as well as the lack of room acoustics within the sound object sounds and the presence of a "buzz tail" at the end of the two mode sound object sounds. As mentioned before, there is a delay of up to 0.3 s when playing an audio file with the Sonic Browser. This delay was judged to be very noticeable when playing sounds within the Sonic Browser. Many participants found this to

Figure 13.17: Representation of the perceptual scaling and tagging information sorted by stimuli.

be irritating and their comments reflected this.

Looking at the results of the three additional questions, questions 6 and 7, which are related to the learnability and interpretation of the application, show that it was easy to learn and understand how to use the application for the task, while question 8 shows that replaying the last sound was judged to be not difficult.

The rich verbal protocol returned several interesting results during the experiment. Many participants found a problem with the scales and on “deciding the scale whether to start with big or small sounds”. Some participants found that it was “much easier to judge size over height”. The “use of speed of repetition as characteristic of height” was found to be helpful in classifying the height of a sound. Another common problem was that the “metallic zips distracts/confuses the classification of sounds”, which refered to the ending of each of the two mode sounds. Another issue illustrated by participants was that a “detailed comparison without reference points is very difficult and would be much easier with only a single scale” and this illustrates the cognitive load of scaling the sounds within a bi-dimensional space. The aura was found to be particularly useful

Figure 13.18: Representation, by a box plot, of the perceptual scaling of the height and size sorted by stimuli.

as “it allows me to see which is higher or which is lower by using pitch. The aura now gives me a comparison for similar sounds”. Another important issue highlighted by participants was that the sound collection consisted of “three divisions (small, medium, large) and that it was very hard to compare between divisions but it was easy to compare within a division”. The divisions refer to the three types of sounds within the space: real sounds, one mode sound objects and two mode sound objects. The participants also spoke about the different materials and surfaces as they found that the “different surfaces are very noticeable”.

One participant (user n.3) performed the task in a very short period compared to the other participants. As told before, he found that “the longer I spent working with the sounds, the more difficult it was to sort them”. This relates to a greater working knowledge of the sound collection and the difficulty in maintaining a consistent scale across multiple sounds. By concentrating on an initial reaction with a continuous exploration and classification of the sound collection it is possible to complete the scaling very quickly but the results showed that quality of the results were only of

Figure 13.19: Representation, by a box plot, of the perceptual scaling of the height and size for stimulus n.7. Uniform perceptual scaling in both the dimensions.

average quality compared to the other participants as shown in Figure 13.16.

13.5 Cataloguing experiment using the Sonic Browser

In the cataloguing scenario we conducted an experiment with a set of tasks to explore common issues arising from the browsing and management of sound object collections. The aim of the experiment is to investigate the usability of the Sonic Browser as a suitable tool for sound object collections and to gain user feedback with regarding the interface which can suggest future improvements or illuminate issues arising from this type of scenario. Specifically this experiment was an exploratory probe designed to further our understanding of managing a large sound collection of sound objects and sound files as well as elaborating upon suitable visual and spatial

Figure 13.20: Representation, by a box plot, of the perceptual scaling of the height and size for stimulus n.11. Uniform perceptual scaling in both the dimensions.

organisations for this type of scenario. In particular, we collected formative data relevant to the understanding of auditory browsing and participant filtering of sounds.

The experiment was preceded by a pilot experiment with two participants. The pilot probe allowed for a preliminary observation of the issues arising from the different type of tasks and it highlighted which types and sound categories were most suitable to focus on in the experiment.

The experiment used the Sonic Browser with three visualisation mechanisms for the visual representations of the sound collection. The three types of visualisation mechanisms used were those of a HyperTree, a TouchGraph and a SObGrid (Starfield display).

A similar experimental setup to the psychophysical experiment was used with both video capture of participants actions and application data logging. The video capture was supplemented by active participant feedback gathered by the Thinking Aloud Protocol as previously described in

Figure 13.21: Representation, by a box plot, of the perceptual scaling of the height and size for stimulus n.10. Slightly uniform perceptual scaling in both the dimensions.

the validation scenario. This type of feedback and commentary from the participants allowed for a greater insight into their perspective of both the tasks and issues encountered within the exploratory probe. The qualitative analysis technique used on the video was a critical incident analysis [164]. The video analysis when used in conjunction with the data logging provides a clear image as to the user actions and intentions in the context of the scenario and each particular task.

In the next section, we will present the cataloguing experiment, introducing both the experiment and the results.

Figure 13.22: Representation, by a box plot, of the perceptual scaling of the height and size for stimulus n.14. Slightly uniform perceptual scaling in both the dimensions.

13.5.1 The experiment

Participants The experiment involved 6 volunteers, all students at the Computer Science Department of the University of Limerick. Five of the participants referred to have musical training in average of 6 years, with a minimum of 4 and a maximum of 12 years. Two of the participants require glasses for reading and one participant reported to have hearing problems with very low tones.

Stimuli The stimuli set included 57 recorded sounds and 10 Sound Objects, designed with pd-modules modelling impact interactions of two modal resonators [9], simplified so as to return only one mode. The recorded sounds used in this experiment were drawn from eight sources: seven commercial sound effects CD's and a local collection of ecological sounds. The length of

Figure 13.23: Representation, by a box plot, of the perceptual scaling of the height and size for stimulus n.17. Highly Spread perceptual scaling.

the sounds varied from 0.1 to 50 seconds. The synthesized sounds used consisted of the same synthesized sounds used in the pilot probe of the validation scenario.

Procedure The experiment was conducted in the isolation room of the Computer Science Department at UL. The sound collection was browsed using stereo headphones and via the Sonic Browser. The experiment was conducted applying the Thinking Aloud Protocol and the participants sessions were video-taped.

The users were asked to browse the sound collection for a selection of sounds matching a particular property or object. The tasks included searching for a specific sound such as the ‘cry of a seagull’ and to broader categories such as find all the sounds of ‘cats meowing’. In each specific task, the participants were allowed to move the cursor around freely in the GUI trying to find target

Figure 13.24: Results of the questionnaire for the main experiment.

Figure 13.25: Results of the questionnaire for the main experiment. These are the data concerning additional questions after the pilot probe.

No.	Question
1	To interpret and understand this design
2	To learn to use this design
3	To find a particular sound, when you know its filename
4	To find a particular sound, when you don't know its filename
5	To find a set of particular sounds with a specific property from a category
6	To perform an AND or an OR query using the filtering mechanism
7	Overall, how would you rate the ease of use or usability of this design
8	Was there a noticeable play lag before a sound was played
9	How realistic did you find the sounds
10	How did you find the quality of the sounds

Table 13.6: Survey questions for the cataloguing experiment.

sounds as well as change the current visualisation at will to compare relationships between sounds within different visualisations. Overall, for the eight auditory tasks, several interesting browsing behaviours were observed. As part of each task, the participants were asked to tag the sounds that they thought fulfilled the criteria of the task.

At the end of each session, as part of the participant debriefing, a questionnaire was presented to the participants in order to gain an insight into their perceptions of the performed tasks and into the Sonic Browser.

Results and Observations In the debriefing phase, a seven point Likert scale questionnaire with six sets of semantic differentials was filled out by the participants who were asked to express their responses to the interfaces and to the tasks, from 0 to 6, where 0 is "poor" and 6 is "excellent". In Table 13.6, we report the survey questions and, in Figure 13.26, the results of the questionnaire with cumulative participant responses displayed per question can be seen with 0 representing a negative result to the question and 6 represent a positive result to the question.

Questions 1, 2 and 7 deal with aesthetics, interpretation and learnability of the Sonic Browser. The results of these questions show that the users find the Sonic Browser easy to learn and use. Questions 3 to 6 deal with the filtering mechanisms of the Sonic Browser. The results of these questions, confirmed by video analysis, illuminate several items such as it is always easier to find a sound when you know its filename and that the filtering mechanisms were found to be easy to use. Question 8 concerns a technical issue which arouse from the piloting phase of this experiment. The results of this question show that in a cataloguing scenario the effect of the delay was not appreciable and did not affect the task. This juxtaposes with the results of this question in the validation scenario which finds that participants had a very noticeable appreciation of the delay. This allows us to say that for tasks involving many sound-to-sound comparisons, play delay of the sound should be kept to an absolute minimum but, in a cataloguing scenario, that, while this is still an important factor, a greater play delay is acceptable. Questions 9 and 10 deal with realism and

Figure 13.26: Results of the questionnaire for the cataloguing experiment.

quality of the sounds which were found to be excellent by participants.

The rich verbal protocol returned several interesting results during the experiment. The play delay was only highlighted by one participant who "expected sound to be instantaneous, not take ages" but this was actually related to the issues of silence at the start of the sound or a sound beginning with a very low volume. The HyperTree visualisation was found to be preferred by half of the participants. Other issues were discovered in the debriefing and through user comments during testing, mostly related to future improvements of the Sonic Browser such as the "Text lookahead should look for 'Flushing toilet' as well as 'Toilet flushing' " and the addition of a "Right click context menu with options for setting object properties and for group of objects properties".

13.6 Conclusions

Examining the results of our validation scenario we can state that synthetic sounds convey information about dimensions even if the resonators are given only one mode. Apart from one case in the pilot probe, the unrealistic perception of sounds did not affect the participant's perception of the sound's dimensions. This illustrates that the "realism" of a sound does not affect the amount of information extracted by a participant. Our studies have shown the difficulty in conveying information in more than one dimension, which is similar to the difficulty encountered with the parameterising of auditory icons [96, 98]. A similar result was shown by Braida's [27] work on multidimensional vibrotactile stimulus which illustrated that, when two dimensions must both be identified, the sensitivity for a particular stimulus is often reduced. Another possible area of difficulty may be due to the orthogonality [139] wherein a change in one parameter of the auditory stream may cause a perception change in another variable. This can arise from a change in pitch or loudness of one of the sound which then affects the perception of the other sounds. Perception distortions can easily be affected by changes to low level acoustic dimensions such as frequency or intensity as discussed by Neuhoff et al [184]. The differing results between the two model parametrizations of sound objects highlight the need for a further investigation of these sound objects exploring any possible perceptual distortions. The results of our study show that unrealistic synthetic sounds can be recognized as unrealistic events but that their high-level parameters can still be extracted and evaluated. This again highlights the technique of sound cartoonification which caricatures certain aspects of a sound event while discarding other aspects. Rath et al. in chapter 9 have already discussed this technique in greater detail with a focus on sound objects in combination with cartoonification.

Another interesting finding was that no sounds were deemed unrealistic by all participants and also that none of the real sounds were selected as being unrealistic. There are several possible explanations for these findings. Our experimental analysis and our user debriefing suggest three aspects which should be further investigated. The first aspect is the inclusion of "room acoustics" and the necessary elements of reverberation within a sound object to allow for a more natural sounding event. The second facet is the material and surface perception by participants which should be further examined as the participants stated that the "different surfaces are very noticeable". The third area is the distractors found within the sounds used in the experiments. The distractors are split into two issues. Firstly, how the participants related the speed and temporal pattern of the bouncing to the realism of the sound. The second is the "metallic zips" occurring at the end of each of the two mode sound objects. These distractors illustrate the need for a further refinement of the perceptual parameters within the sound models to prevent user confusion when judging a sound's properties. Further experiments into the perception of object elasticity and other physical parameters should also be investigated for a greater understanding of the perceptual scaling of Sound Objects.

The cataloguing scenario results allow us to assert that the management of a sound collection is a difficult task but it can be made easier through the use of dynamic filtering combined with both direct manipulation and direct sonification of the sounds within the collection. The Sonic Browser has been successful in providing a new interface which allows for the exploration of a sound collection without the users feeling lost or confused. We plan to continue to enhance the Sonic Browser system building upon our experiences to date as well as those of similar projects. The system and its source will be made available as open-source software and its development will continue. The next stage of this development will be a exploration of the Sonic Browser as a tool for accessing network collections of sound objects, in parallel with the development of greater linkages between pd and the creation of linkage with other audio authoring tools such as open-source tools like Audacity or proprietary tools like SoundForge, Cubase VST or Protools. Even without these planned enhancements, we believe our approach marks a major step forward in audio browsing interfaces and can serve as a model to others implementing similar systems.

Chapter 14

Software tools for Sounding Objects

Nicola Bernardini
Conservatorio di Padova
Padova, Italy
nicb@centrotemporeale.it

14.1 Introduction

In the context of the research described in this book the software development tools and techniques used may be considered of relative importance. However, as it is often the case in research projects, the proper choice of tools and implementation methods can play an essential role in important aspects such as the development of reusable component libraries and demonstration applications.

The requirements set by the Sounding Object research consortium at the beginning of the project were:

1. fast development/debug/testing turnaround time for research and demonstration applications;
2. a coordinated multiple-developer working environment;
3. focus on algorithm development and control mapping coding rather than on nitty-gritty details about audio and control I/O;
4. avoiding proprietary environments and data formats for both ethical and practical reasons;
5. the ability to produce small self-contained applications which may be distributed separately on different platforms.

Given the above requirements, the following tools were adopted:

1. the *Pure Data* (better known as `pd`) musical prototyping application developed by Miller Puckette at the University of California, San Diego [197];
2. the *Concurrent Versions System* service (better known as *CVS* [51]);
3. the *Doxygen* documentation system [118].

The latter two items are widely used in software development and certainly do not need further description here. Therefore, this chapter describes how and to what extent `pd` has satisfied the requirements and allowed the adoption of fast prototyping techniques in software development. At any rate, an important detail of these adoptions is that all tools are licensed under Free Software licenses (GNU/GPL or compatible [228]). All software developed in the Sounding Objects environment is released under the GNU/GPL License as well.

14.2 Rapid prototyping tools

Most Sounding Objects applications and demonstrations have been developed within the `pd` sound environment framework [197]. Essential `pd` features in this context are:

- full client (i.e. GUI)/server (i.e. Sound and control services) system dedicated to sound synthesis, processing and control;
- simple, powerful and flexible messaging system which constitutes the connective backbone among independent software elements;
- fairly advanced plug-in hosting;
- full-blown graphic prototyping (default GUI interface);
- alternative dedicated GUI interfaces (`xgui`, `GriPD`, etc.).

`pd` lacks some features that would be quite useful in the Sounding Objects context, such as:

- musical score management (i.e. the ability of quickly producing polyphonic sound sequences, etc.);
- easy creation of different streaming data types (e.g. spectrum frames, video frames, etc.);
- stand-alone application building.

However, since `pd` is released under a Free Software license, these features can be easily and safely added, and indeed the first two features are being currently added by several contributors. The development of the third feature is a bit more complicated. As it has been designed and created within the Sounding Objects framework, it is described below in section 14.3.

Figure 14.1: Functional `pd` scheme.

Figure 14.1 describes how `pd` works. The default GUI communicates to the server engine the flow-graph of all audio and control paths (i.e.: how single modules and objects are connected together). The server engine provides all hardware management and additional connectivity (MIDI, other interfaces, etc.) and handles them according to whatever has been designed through the used of the GUI application.

`pd` plug-ins sport the following features:

- there is no functional difference between built-in modules and plug-ins other than the run-time initialization dynamic loader;
- as such, all plug-ins access all sound engine and GUI services as the built-in modules do;
- the plug-in structure and API is extremely simple to implement.

Figure 14.2 shows the structure of a `pd` plug-in. Any plug-in designer needs, in principle, to define the following functions for each module:

Creation: Creates the module class, registering its data structure within the system (once per module type).

Run-time Initialization: This function gets called when the module gets instantiated at run-time (for each copy of the module).

Run-time Destruction: This function gets called when the module gets removed from a canvas (for each copy of the module).

Figure 14.2: pd callbacks for each plug-in.

Message Callback Methods: These are functions that are called response to specific messages that the plug-in designer decides to cater.

As in all OOP models, each of these functions gets passed the data structure (i.e. the object) it refers to. In addition, the modules that are used at audio rate need to define another callback which does not get passed the instantiated data structure but rather the signal buffers which need to be treated. The size of the buffers determine the control rate at which this function gets called.

Among the problems that have been encountered in designing physical-model plug-ins for pd in the course of the Sounding Object project, prominent issues have been:

1. the large number of initial condition parameters to be passed to each model at run-time initialization;

2. the need of higher-level algorithmic behavior control for low-level audio synthesis modules.

The first problem arises from the need to keep the models flexible and general enough to encompass a large panoply of different cases (such as *impact models*, *friction models*, etc.). On the other hand, the facility offered by pd to cater these needs proves to be inadequate: pd provides a run-time parameter initialization mechanism by simply passing an ordered sequence of numbers and strings to each module writing them out to the side of the module name. There is no possibility of symbolic binding because this facility has been designed to be used with 2-3 parameters maximum (still, most of pd documentation concerns initialization parameters). When the parameters are in the range of 20-30 (as it is the case for physical modeling modules, this facility proves to be a

Low-level Audio Signal Modules	
Category	**Module**

Impact	`impact_modalb~`
	`impact_2modalb~`
	`linpact_modalb~`
	`linpact_2modalb~`

Friction	`friction_2modalb~`
Rolling	`circ_max_filter~` `decay_max_filter~`

Higher-level Audio Signal Modules	
Category	**Module**
Crumpling	`control_crump`
Walking	`control_footsteps`

Miscellaneous Utilities	
Category	**Module**
Generic Tools	`shell` `clip_exp~`

Table 14.1: `pd` SOb module classification, according to [199].

nightmare to use and maintain even for the most expert sound designers. Of course, defaults can be used. However, these defeat the flexibility and generality purposes of these modules; furthermore, the ordered sequence of parameters hardly allow the use of default values at all.

Luckily, the message-passing mechanism of `pd` is powerful enough to provide a solution to this problem. Other, higher-level specialized plug-ins can be designed to pass complex and sophisticated message sequences to low-level sound synthesis modules in order to provide consistent and precise initialization parameters that are needed to perform specific tasks. These control modules can also take into consideration high-level perceptual features while mapping them into specific low-level parameters.

This double-module structure can also solve the second problem: that is, the need to provide specific algorithmic controls (such as *bouncing*, *rolling*, etc.) to confer the familiar behavior we expect from a realistic auditory landscape[1] [206, 9, 10].

In this framework, the Sounding Object consortium has developed a consistent number of plug-in modules to accomplish a variety of tasks. While these modules continue to be developed, at the time of this writing they can be classified as in Table 14.1 [199]:

[1] this is also true when we approach *cartoonification* issues: high-level behavior is often what allows the recognition of a sound process even in the most abstract situations.

14.3 Self-contained, stand-alone applications

The pd environment has been designed as a fast-prototyping sound development environment. As such, it is hardly suited to develop self-contained stand-alone applications: that is, applications which may not need the user to have installed or to be aware at all of:

a. the pd application itself;

b. all the needed plug-ins to perform a given task;

c. the patch file provided;

d. the dynamically linked libraries of the working environment.

While it is clear that pd was not designed to provide these kind of applications and that trying to get it to produce them is a design abuse, it is also evident that these applications can be very useful in many instances. The Sounding Object project provides at least one of them: easy-to-use stand-alone demonstration patches downloadable from the Internet or distributed in CD-ROMs would indeed be highly appreciated by the scientific and musical community at large. For this reason, we started to design pdplay[2]. pdplay is not an application in itself: rather, it is a set of code files, tools and procedures to produce self-contained stand-alone pd patch applications.

While apparently simple on its surface, the problem of creating self-contained stand-alone applications from straightforward pd patches has proven to be quite difficult. Ideally, pdplay should:

1. use unmodified pd sources as much as possible (to avoid front- and back-porting of an evolving target like pd);

2. use all the facilities provided by pd (nested patches, abstractions, plug-ins, alternative GUIs, etc.).

However, the simplest case already offers a complicate scenario. The bare minimum pd working environment is described in Figure 14.3. This really simple case already shows that a stand-alone application would need to integrate in one multiple-file source:

- the pd source code;
- the tcl/tk script pd.tk;
- the pd patch meant to be the application.

These components are outlined by the dashed rectangle in Figure 14.3.

Figure 14.3: Bare minimum pd setup.

Figure 14.4: A screenshot of pdplay.

This integration has already been done in the current version of pdplay, along with some modifications of the GUI to make it more functional to a stand-alone functionality. Figure 14.4 is

Figure 14.5: An extended pd setup.

a screenshot of what pdplay looks like at the time of this writing.

However, real-world situations are much more intricate when using pd to its fullest extent. A normal user would of course want to use the plug-in facilities of pd, along with carefully crafted sub-patching and perhaps alternative, well-designed dedicated GUIs. Therefore, a (fairly) normal pd environment setup would be similar to that described in Figure 14.5 rather than that show in Figure 14.3. Such a setup poses a number of problems that still need to be solved. Work that needs to be done yet concerns, among other things:

- proper incorporation of the stacking mechanism that allow pd patches to embed other patches or else to include external ones (abstractions);
- proper transformation of dynamic loading of plug-ins into statically linked items;
- easy GUI substitution.

[2] The name was modeled without any fantasy whatsoever from the equivalent proprietary tool *MaxPlay*.

Furthermore, the dynamic linking of libraries poses problems too in stand-alone applications that may be available over the Internet: the quantity of libraries needed and the sheer variety of versions of these libraries spread around make it practically impossible for such stand-alone applications to work on systems other than that which has generated them. The easiest solution would be to use static linking; however, if dynamically-linked `pdplay` applications run up to the 1.8-2.0 Mbytes size it is conceivable that statically linked `pdplay` applications may size up into the 5-6 Mbytes figures. Therefore, other solutions must be conceived: these could range from the use of static but compressed executables to a mixed run-time setup where a common library can be shared among several `pdplay` applications, to an accurate selection of functions that are effectively used in each application. At any rate, these solutions still have to be implemented and tested at the time of this writing.

GNU Free Documentation License

Version 1.2, November 2002

0. PREAMBLE

The purpose of this License is to make a manual, textbook, or other functional and useful document “free” in the sense of freedom: to assure everyone the effective freedom to copy and redistribute it, with or without modifying it, either commercially or noncommercially. Secondarily, this License preserves for the author and publisher a way to get credit for their work, while not being considered responsible for modifications made by others.

This License is a kind of “copyleft”, which means that derivative works of the document must themselves be free in the same sense. It complements the GNU General Public License, which is a copyleft license designed for free software.

We have designed this License in order to use it for manuals for free software, because free software needs free documentation: a free program should come with manuals providing the same freedoms that the software does. But this License is not limited to software manuals; it can be used for any textual work, regardless of subject matter or whether it is published as a printed book. We recommend this License principally for works whose purpose is instruction or reference.

1. APPLICABILITY AND DEFINITIONS

This License applies to any manual or other work, in any medium, that contains a notice placed by the copyright holder saying it can be distributed under the terms of this License. Such a notice grants a world-wide, royalty-free license, unlimited in duration, to use that work under the conditions stated herein. The “Document”, below, refers to any such manual or work. Any member of the public is a licensee, and is addressed as "you". You accept the license if you copy, modify or distribute the work in a way requiring permission under copyright law.

A "Modified Version" of the Document means any work containing the Document or a portion of it, either copied verbatim, or with modifications and/or translated into another language.

A "Secondary Section" is a named appendix or a front-matter section of the Document that deals exclusively with the relationship of the publishers or authors of the Document to the Document's overall subject (or to related matters) and contains nothing that could fall directly within that overall subject. (Thus, if the Document is in part a textbook of mathematics, a Secondary Section may not explain any mathematics.) The relationship could be a matter of historical connection with the subject or with related matters, or of legal, commercial, philosophical, ethical or political position regarding them.

The "Invariant Sections" are certain Secondary Sections whose titles are designated, as being those of Invariant Sections, in the notice that says that the Document is released under this License. If a section does not fit the above definition of Secondary then it is not allowed to be designated as Invariant. The Document may contain zero Invariant Sections. If the Document does not identify any Invariant Sections then there are none.

The "Cover Texts" are certain short passages of text that are listed, as Front-Cover Texts or Back-Cover Texts, in the notice that says that the Document is released under this License. A Front-Cover Text may be at most 5 words, and a Back-Cover Text may be at most 25 words.

A "Transparent" copy of the Document means a machine-readable copy, represented in a format whose specification is available to the general public, that is suitable for revising the document straightforwardly with generic text editors or (for images composed of pixels) generic paint programs or (for drawings) some widely available drawing editor, and that is suitable for input to text formatters or for automatic translation to a variety of formats suitable for input to text formatters. A copy made in an otherwise Transparent file format whose markup, or absence of markup, has been arranged to thwart or discourage subsequent modification by readers is not Transparent. An image format is not Transparent if used for any substantial amount of text. A copy that is not "Transparent" is called "Opaque".

Examples of suitable formats for Transparent copies include plain ASCII without markup, Texinfo input format, LaTeX input format, SGML or XML using a publicly available DTD, and standard-conforming simple HTML, PostScript or PDF designed for human modification. Examples of transparent image formats include PNG, XCF and JPG. Opaque formats include proprietary formats that can be read and edited only by proprietary word processors, SGML or XML for which the DTD and/or processing tools are not generally available, and the machine-generated HTML, PostScript or PDF produced by some word processors for output purposes only.

The "Title Page" means, for a printed book, the title page itself, plus such following pages as are needed to hold, legibly, the material this License requires to appear in the title page. For works in formats which do not have any title page as such, "Title Page" means the text near the most prominent appearance of the work's title, preceding the beginning of the body of the text.

A section "Entitled XYZ" means a named subunit of the Document whose title either is precisely XYZ or contains XYZ in parentheses following text that translates XYZ in another language.

(Here XYZ stands for a specific section name mentioned below, such as “Acknowledgements”, “Dedications”, “Endorsements”, or “History”.) To “Preserve the Title” of such a section when you modify the Document means that it remains a section “Entitled XYZ” according to this definition.

The Document may include Warranty Disclaimers next to the notice which states that this License applies to the Document. These Warranty Disclaimers are considered to be included by reference in this License, but only as regards disclaiming warranties: any other implication that these Warranty Disclaimers may have is void and has no effect on the meaning of this License.

2. VERBATIM COPYING

You may copy and distribute the Document in any medium, either commercially or noncommercially, provided that this License, the copyright notices, and the license notice saying this License applies to the Document are reproduced in all copies, and that you add no other conditions whatsoever to those of this License. You may not use technical measures to obstruct or control the reading or further copying of the copies you make or distribute. However, you may accept compensation in exchange for copies. If you distribute a large enough number of copies you must also follow the conditions in section 3.

You may also lend copies, under the same conditions stated above, and you may publicly display copies.

3. COPYING IN QUANTITY

If you publish printed copies (or copies in media that commonly have printed covers) of the Document, numbering more than 100, and the Document’s license notice requires Cover Texts, you must enclose the copies in covers that carry, clearly and legibly, all these Cover Texts: Front-Cover Texts on the front cover, and Back-Cover Texts on the back cover. Both covers must also clearly and legibly identify you as the publisher of these copies. The front cover must present the full title with all words of the title equally prominent and visible. You may add other material on the covers in addition. Copying with changes limited to the covers, as long as they preserve the title of the Document and satisfy these conditions, can be treated as verbatim copying in other respects.

If the required texts for either cover are too voluminous to fit legibly, you should put the first ones listed (as many as fit reasonably) on the actual cover, and continue the rest onto adjacent pages.

If you publish or distribute Opaque copies of the Document numbering more than 100, you must either include a machine-readable Transparent copy along with each Opaque copy, or state in or with each Opaque copy a computer-network location from which the general network-using public has access to download using public-standard network protocols a complete Transparent copy of the Document, free of added material. If you use the latter option, you must take reasonably prudent steps, when you begin distribution of Opaque copies in quantity, to ensure that this Transparent copy will remain thus accessible at the stated location until at least one year after the last time you distribute an Opaque copy (directly or through your agents or retailers) of that edition to the public.

It is requested, but not required, that you contact the authors of the Document well before redistributing any large number of copies, to give them a chance to provide you with an updated version of the Document.

4. MODIFICATIONS

You may copy and distribute a Modified Version of the Document under the conditions of sections 2 and 3 above, provided that you release the Modified Version under precisely this License, with the Modified Version filling the role of the Document, thus licensing distribution and modification of the Modified Version to whoever possesses a copy of it. In addition, you must do these things in the Modified Version:

A. Use in the Title Page (and on the covers, if any) a title distinct from that of the Document, and from those of previous versions (which should, if there were any, be listed in the History section of the Document). You may use the same title as a previous version if the original publisher of that version gives permission.

B. List on the Title Page, as authors, one or more persons or entities responsible for authorship of the modifications in the Modified Version, together with at least five of the principal authors of the Document (all of its principal authors, if it has fewer than five), unless they release you from this requirement.

C. State on the Title page the name of the publisher of the Modified Version, as the publisher.

D. Preserve all the copyright notices of the Document.

E. Add an appropriate copyright notice for your modifications adjacent to the other copyright notices.

F. Include, immediately after the copyright notices, a license notice giving the public permission to use the Modified Version under the terms of this License, in the form shown in the Addendum below.

G. Preserve in that license notice the full lists of Invariant Sections and required Cover Texts given in the Document's license notice.

H. Include an unaltered copy of this License.

I. Preserve the section Entitled "History", Preserve its Title, and add to it an item stating at least the title, year, new authors, and publisher of the Modified Version as given on the Title Page. If there is no section Entitled "History" in the Document, create one stating the title, year, authors, and publisher of the Document as given on its Title Page, then add an item describing the Modified Version as stated in the previous sentence.

J. Preserve the network location, if any, given in the Document for public access to a Transparent copy of the Document, and likewise the network locations given in the Document for previous versions it was based on. These may be placed in the "History" section. You may omit a network location for a work that was published at least four years before the Document itself, or if the original publisher of the version it refers to gives permission.

K. For any section Entitled "Acknowledgements" or "Dedications", Preserve the Title of the section, and preserve in the section all the substance and tone of each of the contributor acknowledgements and/or dedications given therein.

L. Preserve all the Invariant Sections of the Document, unaltered in their text and in their titles. Section numbers or the equivalent are not considered part of the section titles.

M. Delete any section Entitled "Endorsements". Such a section may not be included in the Modified Version.

N. Do not retitle any existing section to be Entitled "Endorsements" or to conflict in title with any Invariant Section.

O. Preserve any Warranty Disclaimers.

If the Modified Version includes new front-matter sections or appendices that qualify as Secondary Sections and contain no material copied from the Document, you may at your option designate some or all of these sections as invariant. To do this, add their titles to the list of Invariant Sections in the Modified Version's license notice. These titles must be distinct from any other section titles.

You may add a section Entitled "Endorsements", provided it contains nothing but endorsements of your Modified Version by various parties–for example, statements of peer review or that the text has been approved by an organization as the authoritative definition of a standard.

You may add a passage of up to five words as a Front-Cover Text, and a passage of up to 25 words as a Back-Cover Text, to the end of the list of Cover Texts in the Modified Version. Only one passage of Front-Cover Text and one of Back-Cover Text may be added by (or through arrangements made by) any one entity. If the Document already includes a cover text for the same cover, previously added by you or by arrangement made by the same entity you are acting on behalf of, you may not add another; but you may replace the old one, on explicit permission from the previous publisher that added the old one.

The author(s) and publisher(s) of the Document do not by this License give permission to use their names for publicity for or to assert or imply endorsement of any Modified Version.

5. COMBINING DOCUMENTS

You may combine the Document with other documents released under this License, under the terms defined in section 4 above for modified versions, provided that you include in the combination all of the Invariant Sections of all of the original documents, unmodified, and list them all as Invariant Sections of your combined work in its license notice, and that you preserve all their Warranty Disclaimers.

The combined work need only contain one copy of this License, and multiple identical Invariant Sections may be replaced with a single copy. If there are multiple Invariant Sections with the same name but different contents, make the title of each such section unique by adding at the end of it, in parentheses, the name of the original author or publisher of that section if known, or else a unique number. Make the same adjustment to the section titles in the list of Invariant Sections in the license notice of the combined work.

In the combination, you must combine any sections Entitled “History” in the various original documents, forming one section Entitled “History”; likewise combine any sections Entitled “Acknowledgements”, and any sections Entitled “Dedications”. You must delete all sections Entitled “Endorsements”.

6. COLLECTIONS OF DOCUMENTS

You may make a collection consisting of the Document and other documents released under this License, and replace the individual copies of this License in the various documents with a single copy that is included in the collection, provided that you follow the rules of this License for verbatim copying of each of the documents in all other respects.

You may extract a single document from such a collection, and distribute it individually under this License, provided you insert a copy of this License into the extracted document, and follow this License in all other respects regarding verbatim copying of that document.

7. AGGREGATION WITH INDEPENDENT WORKS

A compilation of the Document or its derivatives with other separate and independent documents or works, in or on a volume of a storage or distribution medium, is called an “aggregate” if the copyright resulting from the compilation is not used to limit the legal rights of the compilation’s users beyond what the individual works permit. When the Document is included in an aggregate, this License does not apply to the other works in the aggregate which are not themselves derivative works of the Document.

If the Cover Text requirement of section 3 is applicable to these copies of the Document, then if the Document is less than one half of the entire aggregate, the Document’s Cover Texts may be placed on covers that bracket the Document within the aggregate, or the electronic equivalent of covers if the Document is in electronic form. Otherwise they must appear on printed covers that bracket the whole aggregate.

8. TRANSLATION

Translation is considered a kind of modification, so you may distribute translations of the Document under the terms of section 4. Replacing Invariant Sections with translations requires special permission from their copyright holders, but you may include translations of some or all Invariant Sections in addition to the original versions of these Invariant Sections. You may include a translation of this License, and all the license notices in the Document, and any Warranty Disclaimers, provided that you also include the original English version of this License and the original versions of those notices and disclaimers. In case of a disagreement between the translation and the original version of this License or a notice or disclaimer, the original version will prevail.

If a section in the Document is Entitled “Acknowledgements”, “Dedications”, or “History”, the requirement (section 4) to Preserve its Title (section 1) will typically require changing the actual title.

9. TERMINATION

You may not copy, modify, sublicense, or distribute the Document except as expressly provided for under this License. Any other attempt to copy, modify, sublicense or distribute the Document is void, and will automatically terminate your rights under this License. However, parties who have received copies, or rights, from you under this License will not have their licenses terminated so long as such parties remain in full compliance.

10. FUTURE REVISIONS OF THIS LICENSE

The Free Software Foundation may publish new, revised versions of the GNU Free Documentation License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns. See http://www.gnu.org/copyleft/.

Each version of the License is given a distinguishing version number. If the Document specifies that a particular numbered version of this License “or any later version” applies to it, you have the option of following the terms and conditions either of that specified version or of any later version that has been published (not as a draft) by the Free Software Foundation. If the Document does not specify a version number of this License, you may choose any version ever published (not as a draft) by the Free Software Foundation.

ADDENDUM: How to use this License for your documents

To use this License in a document you have written, include a copy of the License in the document and put the following copyright and license notices just after the title page:

> Copyright (c) YEAR YOUR NAME. Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.2 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts. A copy of the license is included in the section entitled “GNU Free Documentation License”.

If you have Invariant Sections, Front-Cover Texts and Back-Cover Texts, replace the “with...Texts.” line with this:

> with the Invariant Sections being LIST THEIR TITLES, with the Front-Cover Texts being LIST, and with the Back-Cover Texts being LIST.

If you have Invariant Sections without Cover Texts, or some other combination of the three, merge those two alternatives to suit the situation.

If your document contains nontrivial examples of program code, we recommend releasing these examples in parallel under your choice of free software license, such as the GNU General Public License, to permit their use in free software.

References

[1] Adc-11. Pico Technology Limited, registered trademark, http://www.picotech.com.

[2] J. M. Adrien. The missing link: Modal synthesis. In G. De Poli, A. Piccialli, and C. Roads, editors, *Representations of Musical Signals*, pages 269–297. MIT Press, Cambridge, MA, 1991.

[3] C. Ahlberg and B. Shneiderman. The Alphaslider: A rapid and compact selector. In *Proc. ACM CHI*, pages 365–371, Boston, MA, 1994.

[4] C. Ahlberg and B. Shneiderman. Visual information seeking: Tight coupling of dynamic query filters with starfield displays. In *Proc. ACM CHI*, pages 313–317, Boston, MA, 1994.

[5] A. Akay. Acoustics of friction. *J. of the Acoustical Society of America*, 111(4):1525–1548, 2002.

[6] M. C. Albers. The Varese system, hybrid auditory interfaces, and satellite-ground control: Using auditory icons and sonification in a complex, supervisory control system. In G. Kramer and S. Smith, editors, *Proc. Int. Conf. on Auditory Display*, pages 3–14, Santa Fe, NM, November 1994.

[7] F. Altpeter. *Friction Modeling, Identification and Compensation.* PhD thesis, École Polytechnique Fédérale de Lausanne, 1999.

[8] F. Avanzini. *Computational issues in physically-based sound models*. PhD thesis, Università degli Studi di Padova, 2001.

[9] F. Avanzini, M. Rath, and D. Rocchesso. Physically–based audio rendering of contact. In *Proc. IEEE Int. Conf. on Multimedia & Expo*, Lausanne, August 2002.

[10] F. Avanzini and D. Rocchesso. Controlling material properties in physical models of sounding objects. In *Proc. Int. Computer Music Conference*, La Habana, Cuba, September 2001.

[11] F. Avanzini and D. Rocchesso. Modeling Collision Sounds: Non-linear Contact Force. In *Proc. Conf. on Digital Audio Effects*, pages 61–66, Limerick, December 2001.

[12] F. Avanzini and D. Rocchesso. Efficiency, accuracy, and stability issues in discrete time simulations of single reed instruments. *J. of the Acoustical Society of America*, 111(5):2293–2301, May 2002.

[13] J. A. Ballas. Common factors in the identification of an assortment of brief everyday sounds. *J. of Experimental Psychology: Human Perception and Performance*, 19(2):250–267, 1993.

[14] J. A. Ballas and J. H. Howard. Interpreting the language of environmental sounds. *Environment and Behaviour*, 9(1):91–114, 1987.

[15] A. Barr. Superquadrics and angle-preserving transforms. *IEEE Computer Graphics and Applications*, 1(1):11–23, 1981.

[16] S. Barrass. *Auditory Information Design*. PhD thesis, Australian National University, 1997.

[17] L. L. Beranek. Concert hall acoustics - 1992. *J. of the Acoustical Society of America*, 92(1):1–39, 1992.

[18] L. Berio and N. Bernardini. Incontro con Luciano Berio. In *La passione del conoscere*. AA.VV., Laterza Ed., Bari, Italy, 1993.

[19] M. M. Blattner. Investigations into the use of earcons. Technical report, Lawrence Livermore National Laboratory, University of California, Davis, CA, 1996.

[20] M. M. Blattner, D. Sumikawa, and R. Greenberg. Earcons and icons: Their structure and common design principles. *Human-Computer Interaction*, 4(1), Spring 1989.

[21] J. Blauert. *Spatial Hearing: the Psychophysics of Human Sound Localization*. MIT Press, Cambridge, MA, 1983.

[22] M. T. Boren and J. Ramey. Thinking aloud: Reconciling theory and practice. *IEEE Trans. on Professional Communication*, 43(3):261–278, September 2000.

[23] G. Borin and G. De Poli. A hysteretic hammer-string interaction model for physical model synthesis. In *Proc. Nordic Acoustical Meeting*, pages 399–406, Helsinki, June 1996.

[24] G. Borin, G. De Poli, and D. Rocchesso. Elimination of delay-free loops in discrete-time models of nonlinear acoustic systems. *IEEE Trans. on Speech and Audio Processing*, 8(5):597–605, 2000.

[25] R. Boulanger and M. Mathews. The 1997 Mathews radio-baton and improvisation modes. In *Proc. Int. Computer Music Conference*, pages 395–398, 1997.

[26] L. H. Boyd, W. L. Boyd, and G. C. Vanderheiden. The graphical user interface: Crisis, danger, and opportunity. *Visual Impairment & Blindness*, pages 496–502, 1990.

[27] L. D. Braida. Development of a model for multidimensional identification experiments. *J. of the Acoustical Society of America*, 87, 1988.

[28] E. Brazil and M. Fernström. Let your ears do the browsing - the Sonic Browser. *The Irish Scientist*, 2001.

[29] E. Brazil, M. Fernström, G. Tzanetakis, and P. Cook. Enhancing sonic browsing using audio information retrieval. In *Proc. Int. Conf. on Auditory Display*, Kyoto, Japan, 2002.

[30] A. Bregman. *Auditory Scene Analysis*. The MIT Press, Cambridge, MA, 1990.

[31] R. Bresin. Articulation rules for automatic music performance. In *Proc. Int. Computer Music Conference*, La Habana, Cuba, September 2001.

[32] R. Bresin and G. U. Battel. Articulation strategies in expressive piano performance. Analysis of legato, staccato, and repeated notes in performances of the andante movement of Mozart's sonata in G major (K 545). *J. of New Music Research*, 29(3):211–224, 2000.

[33] R. Bresin and A. Friberg. Emotional coloring of computer-controlled music performances. *Computer Music J.*, 24(4):44–63, 2000. MIT press.

[34] R. Bresin and G. Widmer. Production of staccato articulation in mozart sonatas played on a grand piano. Preliminary results. Speech Music and Hearing Quarterly Progress and Status Report 4, Speech Music and Hearing, Stockholm: KTH, 2000.

[35] S. A. Brewster. Sound in the interface to a mobile computer. In Lawrence Erlbaum Associates, editor, *Proc. of HCI International*, pages 43–47, Munich, Germany, 1999.

[36] S. A. Brewster. Overcoming the lack of screen space on mobile computers. *Personal and Ubiquitous Computing*, 6(3):188–205, 2002.

[37] S. A. Brewster, P. C. Wright, A. J. Dix, and A. D. N. Edwards. The sonic enhancement of graphical buttons. In K. Nordby, P. Helmersen, D. Gilmore, and S. Arnesen, editors, *Proc. of Interact'95*, pages 43–48, Lillehammer, Norway, 1995. Chapman & Hall.

[38] S. A. Brewster, P. C. Wright, and A. D. N. Edwards. The design and evaluation of an auditory-enhanced scrollbar. In S. Dumais B. Adelson and J. Olson, editors, *Proc. ACM CHI*, pages 173–179, Boston, MA, 1994. ACM Press, Addison-Wesley.

[39] S. A. Brewster, P. C. Wright, and A. D. N. Edwards. A detailed investigation into the effectiveness of earcons. In G. Kramer, editor, *Auditory Display: Sonification, Audification and Auditory interfaces*, pages 471–498. Addison-Wesley, Reading, MA, 1994.

[40] D. S. Brungart, N. I. Durlach, and W. M. Rabinowitz. Auditory localization of nearby sources. II. localization of a broadband source. *J. of the Acoustical Society of America*, 106:1956–1968, 1999.

[41] Mr Bungle. Mr Bungle. CD, 1991: 2-26640. New York: Warner.

[42] W. Buxton and W. W. Gaver. Human interface design and the handicapped user. In *ACM CHI'86*. ACM Press, 1986.

[43] W. Buxton and W. W. Gaver. The use of non-speech audio at the interface. In *CHI'89*, Austin, Texas, 1989. ACM Press.

[44] W. Buxton, W. W. Gaver, and S. Bly. Non-speech audio at the interface. Unfinished book manuscript, http://www.billbuxton.com/Audio.TOC.html, 1994.

[45] D. Byrne. Feelings®. CD, 1997: 46605. New York: Luaka Bop/Warner Brothers.

[46] P. A. Cabe and J. B. Pittenger. Human sensitivity to acoustic information from vessel filing. *J. of Experimental Psychology: Human Perception and Performance*, 26(1):313–324, 2000.

[47] J. Cage. Imaginary landscape no 1, 1939. Edition Peters - New York.

[48] J. W. Cannon, W. J. Floyd, R. Kenyon, and W. R. Parry. Hyperbolic geometry. In S. Levy, editor, *Flavors of Geometry*, pages 59–115. Cambridge University Press, Cambridge, 1997.

[49] C. Carello, K. L. Anderson, and A. J. Kunkler-Peck. Perception of object length by sound. *Psychological Science*, 9(3):211–214, May 1998.

[50] C. Carello, J. B. Wagman, and M. T. Turvey. Acoustical specification of object properties. In J. Anderson and B. Anderson, editors, *Moving image theory: Ecological considerations*. Southern Illinois University Press, Carbondale, IL, 2003.

[51] P. Cederqvist. *Version Management with CVS*, 2003. Version 11.3 web available at http://www.cvshome.org/docs/manual.

[52] A. Chaigne and C. Lambourg. Time-domain simulation of damped impacted plates: I. theory and experiments. *J. of the Acoustical Society of America*, 109(4):1422–1432, April 2001.

[53] A. Chaigne, C. Lambourg, and D. Matignon. Time-domain simulation of damped impacted plates: II. numerical models and results. *J. of the Acoustical Society of America*, 109(4):1433–1447, April 2001.

[54] P. R. Cook. *Real Sound Synthesis for Interactive Applications*. A. K. Peters Ltd., 2002.

[55] G. Corsini and R. Saletti. A $1/f^\gamma$ power spectrum noise sequence generator. *IEEE Trans. on Instrumentation and Measurement*, 37(4):615–619, December 1988.

[56] C. Cutler. Plunderphonia. http://www.l-m-c.org.uk/texts/plunder.html, 1994. MusicWorks 60, 6-20. Toronto: MusicWorks.

[57] S. Dahl. Spectral changes in the tom-tom related to striking force. Speech Music and Hearing Quarterly Progress and Status Report 1, Dept. of Speech Music and Hearing, Stockholm: KTH, 1997.

[58] S. Dahl. The playing of an accent. preliminary observations from temporal and kinematic analysis of percussionists. *J. of New Music Research*, 29(3):225–234, 2000.

[59] S. Dahl and R. Bresin. Is the player more influenced by the auditory than the tactile feedback from the instrument? In *Proc. Conf. on Digital Audio Effects*, pages 194–197, Limerick, 2001.

[60] R. Dannenberg and I. Derenyi. Combining instrument and performance models for high-quality music synthesis. *J. of New Music Research*, 27(3):211–238, 1998.

[61] ddrum4. Registered trademark of Clavia Digital Musical Instruments AB, http://www.clavia.se/ddrum.htm.

[62] C. Canudas de Wit, C. H. Olsson, K. J. Åström, and P. Lischinsky. A new model for control of systems with friction. *IEEE Trans. Autom. Control*, 40(3):419–425, 1995.

[63] D. Deutsch. *The Psychology of Music*. Academic Press, San Diego, 1999.

[64] Hitech Development. Soundswell signal workstation . http://www.hitech.se/development/products/soundswell.htm.

[65] P. Djoharian. Shape and material design in physical modeling sound synthesis. In *Proc. Int. Computer Music Conference*, pages 38–45, Berlin, 2000.

[66] P. Dupont, V. Hayward, B. Armstrong, and F. Altpeter. Single State Elasto-Plastic Friction Models. *IEEE Trans. Autom. Control*, (5):787–792, 2002.

[67] N.I. Durlach, B.G. Shinn-Cunningham, and R.M. Held. Supernormal auditory localization. I. General background. *Presence: Teleoperators and Virtual Environment*, 2(2):89–103, 1993.

[68] S. A. Van Duyne and J. O. Smith. A simplified approach to modeling dispersion caused by stiffness in strings and plates. In *Proc. Int. Computer Music Conference*, pages 407–410, Aarhus, Denmark, September 1994.

[69] S.A. Van Duyne and J.O. Smith. Physical modeling with the 2-D digital waveguide mesh. In *Proc. Int. Computer Music Conference*, pages 40–47, Tokyo, Japan, 1993.

[70] T. Engen. Psychophysics. I. Discrimination and detection. II. Scaling. In *Woodworth & Schlosberg's Experimental Psychology*, pages 11–86. J.K. Kling and L.A. Riggs (ed.), Methuen, London, 3rd edition, 1971.

[71] M. Fernström and C. McNamara. After direct manipulation - direct sonification. In *Proc. Int. Conf. on Auditory Display*, Glasgow, Scotland, 1998.

[72] S. A. Finney. Auditory feedback and musical keyboard performance. *Music Perception*, 15(2):153–174, 1997.

[73] W. T. Fitch and G. Kramer. Sonifying the body electric: Superiority of an auditory display over a visual display in a complex, multivariate system. In G. Kramer, editor, *Auditory Display: Sonification, Audification and Auditory interfaces*, pages 307–326. Addison-Wesley, Reading, MA, 1994.

[74] T. Flash and N. Hogan. The coordination of arm movements: an experimentally confirmed mathematical model. *The J. of Neuroscience*, 5(7):1688–1703, 1985.

[75] H. Fletcher and W. A. Munson. Loudness, its definition measurements and calculation. *J. of the Acoustical Society of America*, 5:82–108, 1933.

[76] N. H. Fletcher and T. D. Rossing. *The Physics of Musical Instruments*. Springer-Verlag, New York, 1991.

[77] F. Fontana and R. Bresin. Physics-based sound synthesis and control: crushing, walking and running by crumpling sounds. In *Proc. Colloquium on Musical Informatics*, Florence, Italy, May 2003.

[78] F. Fontana, D. Rocchesso, and E. Apollonio. Using the waveguide mesh in modelling 3D resonators. In *Proc. Conf. on Digital Audio Effects*, pages 229–232, Verona - Italy, December 2000.

[79] C. A. Fowler. Sound-producing sources of perception: Rate normalization and nonspeech perception. *J. of the Acoustical Society of America*, 88(6):1236–1249, June 1990.

[80] C. A. Fowler. Auditory perception is not special: we see the world, we feel the world, we hear the world. *J. of the Acoustical Society of America*, 89(6):2910–2915, June 1991.

[81] Fab Five Freddie. Change the beat. 12” Celluloid Records CEL 156, 1982.

[82] D. J. Freed. Auditory correlates of perceived mallet hardness for a set of recorded percussive events. *J. of the Acoustical Society of America*, 87(1):311–322, January 1990.

[83] A. Friberg. Generative rules for music performance: A formal description of a rule system. *Computer Music J.*, 15(2):56–71, 1991.

[84] A. Friberg, V. Colombo, L. Frydén, and J. Sundberg. Generating musical performances with director musices. *Computer Music J.*, 24(3):23–29, 2000.

[85] A. Friberg and J. Sundberg. Time discrimination in a monotonic, isochrounous sequence. *J. of the Acoustical Society of America*, 98(5):2524–2531, 1995.

[86] A. Friberg and J. Sundberg. Does music performance allude to locomotion? a model of final ritardandi derived from measurements of stopping runners. *J. of the Acoustical Society of America*, 105(3):1469–1484, 1999.

[87] A. Friberg, J. Sundberg, and L. Frydén. Music from motion: Sound level envelopes of tones expressing human locomotion. *J. of New Music Research*, 29(3):199–210, 2000.

[88] G. Furnas. Generalised fisheye views. In *Proc. ACM CHI*, Massachusetts, USA, 1986.

[89] G. Furnas. Space-scale diagrams: Understanding multiscale interfaces. In *Proc. ACM CHI*, Denver, CO, USA, 1995.

[90] W. G. Gardner. Reverberation algorithms. In M. Kahrs and K. Brandenburg, editors, *Applications of Digital Signal Processing to Audio and Acoustics*, pages 85–131. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1998.

[91] A. Gates, J. Bradshaw, and N. Nettleton. Effect of different delayed auditory feedback intervals on a music performance task. *Perception and Psychophysics*, 14:21–25, 1974.

[92] W. W. Gaver. *Everyday listening and auditory icons*. PhD thesis, University of California, San Diego, 1988.

[93] W. W. Gaver. The sonic finder: An interface that uses auditory icons. the use of non-speech audio at the interface. In *Proc. ACM CHI*, Austin, Texas, 1989.

[94] W. W. Gaver. Sound support for collaboration. In *Proc. of ECSCW'91*, Amsterdam, The Netherlands, September 1991. Kluwer. Reprinted in R. Baecker, (ed.), Readings in groupware and CSCW: Assisting human-human collaboration. Morgan Kaufmann, San Mateo, CA, 1993.

[95] W. W. Gaver. How do we hear in the world? explorations in ecological acoustics. *Ecological Psychology*, 5(4):285–313, 1993.

[96] W. W. Gaver. Synthesizing auditory icons. In *INTERCHI '93, 24-29 April 1993*, pages 228–235, 1993.

[97] W. W. Gaver. What in the world do we hear? an ecological approach to auditory event perception. *Ecological Psychology*, 5(1):1–29, 1993.

[98] W. W. Gaver. Using and creating auditory icons. In G. Kramer, editor, *Auditory Display: Sonification, Audification and Auditory interfaces*, pages 417–446. Addison-Wesley, Reading, MA, 1994.

[99] W. W. Gaver. Auditory interfaces. In M. G. Helander, T. K. Landauer, and P. Prabhu, editors, *Handbook of Human-Computer Interaction*. Elsevier Science, Amsterdam, The Netherlands, 2nd edition, 1997.

[100] W. W. Gaver, R. Smith, and T. O'Shea. Effective sounds in complex systems: the arkola simulation. In *Proc. ACM CHI*, New Orleans, Louisiana, 19991.

[101] J. J. Gibson. A theory of pictorial perception. *Audio-Visual Communication Review*, 1:3–23, 1954.

[102] J. J. Gibson. The useful dimension of sensitivity. *American Psychologist*, 18:115, 1963.

[103] J. J. Gibson. *The ecological approach to visual perception*. Houghton Mifflin, Boston, 1979.

[104] S. Godlovitch. *Musical Performance: A philosophical study*. Routledge, London, UK, 1998.

[105] S. Granqvist. Enhancements to the visual analogue scale, VAS, for listening tests. Speech Music and Hearing Quarterly Progress and Status Report 4, Speech Music and Hearing, Stockholm: KTH, 1996.

[106] J. M. Grey and J. W. Gordon. Perceptual effects of spectral modifications on musical timbres. *J. of the Acoustical Society of America*, 63:1493–1500, 1978.

[107] P. Griffiths. *Modern music and after*. Oxford University Press, 1995.

[108] R. Guski. Auditory localization: effects of reflecting surfaces. *Perception*, 19:819–830, 1990.

[109] B. Gygi. *Factors in the identification of environmental sounds*. PhD thesis, Indiana University, Department of Psychology, 2001.

[110] D. E. Hall. *Musical Acoustics*. Wadsworth, 1980.

[111] D. E. Hall. Piano string excitation VI: Nonlinear modeling. *J. of the Acoustical Society of America*, 92:95–105, July 1992.

[112] H. Hancock. Future shock. CD, reissue 2000: 65962. New York: Columbia/Legacy.

[113] S. Handel. Timbre perception and auditory object identification. In B.C.J. Moore, editor, *Hearing*. Academic Press, San Diego, CA, 1995.

[114] K. F. Hansen. Turntable music. In L. Jonsson, K. Oversand, and M. Breivik, editors, *Musikklidenskapelig Årbok 2000*, pages 145–160. Trondheim: NTNU, 2000. http://www.speech.kth.se/˜hansen/turntablemusic.html.

[115] K.F. Hansen. Turntablisme - his master's voice: The art of the record player. Master's thesis, NTNU, Trondheim: NTNU, 1999.

[116] H. M. Hastings and G. Sugihara. *Fractals: A User's Guide for the Natural Sciences*. Oxford University Press, 1993.

[117] V. Hayward and B. Armstrong. A new computational model of friction applied to haptic rendering. In P. Corke and J. Trevelyan, editors, *Experimental Robotics VI*, pages 403–412. Springer-Verlag, 2000.

[118] D. Van Heesch. *The Doxygen Manual*, 2003. Version 1.3-rc3 web available at http://www.stack.nl/˜dimitri/doxygen/manual.html.

[119] F. Heider and M. Simmel. A study of apparent behavior. *American J. of Psychology*, 57:243–256, 1944.

[120] M. M. J. Houben, A. Kohlrausch, and D. J. Hermes. Auditory cues determining the perception of size and speed of rolling balls. In *Proc. Int. Conf. on Auditory Display*, pages 105–110, Espoo, Finland, 2001.

[121] P. A. Houle and J. P. Sethna. Acoustic emission from crumpling paper. *Physical Review E*, 54(1):278–283, July 1996.

[122] K. H. Hunt and F. R. E. Crossley. Coefficient of Restitution Interpreted as Damping in Vibroimpact. *ASME J. Applied Mech.*, pages 440–445, June 1975.

[123] J. Huopaniemi, L. Savioja, and M. Karjalainen. Modeling of reflections and air absorption in acoustical spaces: a digital filter design approach. In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pages 19–22, Mohonk, NY, 1997. IEEE.

[124] E. L. Hutchins, J. D. Hollan, and D. A. Norman. Direct manipulation interfaces. In D. A. Norman and S. W. Draper, editors, *User Centered System Design: New Perspectives on Human-Computer Interaction*, pages 87–124. Lawrence Erlbaum Associates, Hillsdale, NJ, 1986.

[125] DJ 1210 Jazz. Book of five scratches. Book 2. Snickars Rec., SR1206, 2001.

[126] G. Johansson. Studies on visual perception of locomotion. *Perception*, 6:365–376, 1977.

[127] B. Johnson and B. Shneiderman. Tree-maps: A space filling approach to the visualization of hierarchical information structures. In *IEEE Visualization '91*, San Diego, CA, 1991.

[128] J.-M. Jot. An Analysis/Synthesis Approach to Real-Time Artificial Reverberation. In *Proc. Int. Conf. on Acoustics, Speech, and Signal Processing*, San Francisco, CA, 1992. IEEE.

[129] J.-M. Jot. Real-time spatial processing of sounds for music, multimedia and interactive human-computer interfaces. *Multimedia systems*, 7:55–69, 1999.

[130] W. B. Joyce. Exact effect of surface roughness on the reverberation time of a uniformly absorbing spherical enclosure. *J. of the Acoustical Society of America*, 64(5):1429–1436, 1978.

[131] N.P. Juslin, A. Friberg, and R. Bresin. Toward a computational model of expression in music performance. The GERM model. *Musica Scientiae*, 2002.

[132] G. Kanizsa. Condizioni ed effetti della trasparenza fenomenica [conditions and the effects of perceptual transparency]. *Rivista di Psicologia*, 49:3–19, 1955.

[133] G. Kanizsa and G. B. Vicario. La perception de la réaction intentionnelle. *Bulletin de Psychologie*, 234:1019–1039, 1970.

[134] D. Katz. Die erscheinungsweisen der farben und ihre beeinflussung durch die individuelle erfahrung. *Zeitschrift für Psychologie*, 1911. (special number 7). *The world of color*, Paul Kegan, London 1921.

[135] M. Kennedy. *Oxford Concise Dictionary of Music*. Oxford University Press, Oxford, 1996.

[136] R. Khazam. Electroacoustic alchemist. *The Wire Magazine*, pages 36–40, 1997.

[137] R. L. Klatzky, D. K. Pai, and E. P. Krotkov. Perception of material from contact sounds. *Presence: Teleoperators and Virtual Environment*, 9(4):399–410, August 2000.

[138] K. Koffka. *Principles of Gestalt Psychology*. Routledge & Kegan, London, 1935. (1935/1962 5).

[139] G. Kramer. Some organizing principles for representing data with sound. In G. Kramer, editor, *Auditory Display: Sonification, Audification and Auditory interfaces*, pages 185–222. Addison-Wesley, 1994.

[140] M. Kubovy and D. Van Valkenburg. Auditory and visual objects. *Cognition*, 80:97–126, 2001.

[141] A. Kulkarni and H. S. Colburn. Role of spectral detail in sound-source localization. *Nature*, 396:747–749, December 1998.

[142] A. J. Kunkler-Peck and M. T. Turvey. Hearing shape. *J. of Experimental Psychology: Human Perception and Performance*, 26(1):279–294, 2000.

[143] H. Kuttruff. *Room Acoustics*. Elsevier Science, Essex, England, 1973. Third Edition, 1991.

[144] S. Lakatos. A common perceptual space for harmonic and percussive timbres. *Perception & Psychophysics*, 62:1426–1439, 2000.

[145] S. Lakatos, S. McAdams, and R. Caussé. The representation of auditory source characteristics: simple geometric form. *Perception & Psychophysics*, 59(8):1180–1190, 1997.

[146] J. D. Lambert. *Numerical Methods for Ordinary Differential Systems*. John Wiley & Sons, Chichester, UK, 1993.

[147] J. Lamping, R. Rao, and P. Pirolli. A focus+context technique based on hyperbolic geometry for visualizing large hierarchies. In *Proc. ACM CHI*, Denver, CO, USA, 1995.

[148] S. J. Lederman. Auditory texture perception. *Perception*, 8:93–103, 1979.

[149] D. N. Lee. A theory of visual control of breaking based on information about time-to-collision. *Perception*, 5:437–459, 1976.

[150] G. Leplatre and S. A. Brewster. An investigation of using music to provide navigation cues. In *Proc. Int. Conf. on Auditory Display*, Glasgow, UK, 1998. British Computer Society.

[151] Y. K. Leung and M. D. Apperley. A review and taxonomy of distortion-oriented presentation techniques. *ACM Trans. on Computer-Human Interaction*, 1:126–160, 1994.

[152] H. Levitt. Transformed up-down methods in psychoacoustics. *J. of the Acoustical Society of America*, 49(2):467–477, 1970.

[153] X. Li, R. J. Logan, and R. E. Pastore. Perception of acoustic source characteristics: Walking sounds. *J. of the Acoustical Society of America*, 90(6):3036–3049, December 1991.

[154] A. L. Liberman and I. G. Mattingly. The motor theory of speech perception revised. *Cognition*, 21:1–36, 1985.

[155] B. Libet. Cerebral processes that distinguish conscious experience from unconscious mental functions. In J.C. Eccles and O. Creutsfeldt, editors, *The principles of design and operation of the brain*, pages 185–202. Pontificia Academia Scientiarum, Roma, 1990.

[156] B. Lindblomm. Explaining phonetic variation: a sketch of the h&h theory. In Hardcastle & Marchal, editor, *Speech production and speech modeling*, pages 403–439. Kluwer, Dordrecht, 1990.

[157] J. M. Loomis, R. L. Klatzky, and R. G. Golledge. Auditory distance perception in real, virtual, and mixed environments. In Y. Ohta and H. Tamura, editors, *Mixed Reality: Merging Real and Virtual Worlds*, pages 201–214. Ohmsha, Ltd., Tokio, Japan, 1999.

[158] K. Lorenz. *Die Rückseite des Spiegels. Versuche einer Naturgeschichte menschlichen Erkennens*. Piper & Co., München, 1973.

[159] R. A. Lutfi. Informational processing of complex sound. I: Intensity discrimination. *J. of the Acoustical Society of America*, 86(3):934–944, 1989.

[160] R. A. Lutfi. Informational processing of complex sound. II: Cross-dimensional analysis. *J. of the Acoustical Society of America*, 87(5):2141–2148, 1990.

[161] R. A. Lutfi. Informational processing of complex sound. III: Interference. *J. of the Acoustical Society of America*, 91(6):3391–3401, 1992.

[162] R. A. Lutfi. Auditory detection of hollowness. *J. of the Acoustical Society of America*, 110(2), August 2001.

[163] R. A. Lutfi and E. L. Oh. Auditory discrimination of material changes in a struck-clamped bar. *J. of the Acoustical Society of America*, 102(6):3647–3656, December 1997.

[164] W. E. Mackay. Using video to support interaction design. Technical report, INRIA DISC multimedia, 2002.

[165] C.L. MacKenzie and D.L. Van Erde. Rhythmic precision in the performance of piano scales: Motor psychophysics and motor programming. In M. Jeannerod, editor, *Proc. Int. Symposium on Attention and Performance*, pages 375–408, Hillsdale, 1990. Lawrence Erlbaum Associates.

[166] G. Madison. Drumming performance with and without clicktrack - the validity of the internal clock in expert synchronisation. In *Proc. Fourth Workshop on Rhythm Perception & Production*, pages 117–122, Bourges, France, 1992.

[167] D. W. Marhefka and D. E. Orin. A compliant contact model with nonlinear damping for simulation of robotic systems. *IEEE Trans. on Systems, Man, and Cybernetics - Part A: Systems and Humans*, 29(6):566–572, November 1999.

[168] M. Marshall, M. Rath, and B. Moynihan. The Virtual Bhodran — The Vodhran. In *Proc. Int. Workshop on New Interfaces for Musical Expression*, Dublin, May 2002.

[169] W. L. Martens. Psychophysical calibration for controlling the range of a virtual sound source: Multidimensional complexity in spatial auditory display. In *Proc. Int. Conf. on Auditory Display*, pages 197–207, Espoo, Finland, July 2001.

[170] S. C. Masin. *Foundations of perceptual theory*. North-Holland, Amsterdam, 1993.

[171] S. McAdams. Recognition of sound sources and events. In S. McAdams and E. Bigand, editors, *Thinking in sound: the cognitive psychology of human audition*, pages 146–198. Oxford University Press, 1993.

[172] R. McGrath, T. Waldmann, and M. Fernström. Listening to rooms and objects. In *Proc. AES Conf. on Spatial Sound Reproduction*, pages 512–522, Rovaniemi, Finland, April 1999.

[173] M. Mellody and G. H. Wakefield. A tutorial example of Stimulus Sample Discrimination in perceptual evaluation of synthesized sounds: discrimination between original and re-synthesized singing. In *Proc. Int. Conf. on Auditory Display*, Espoo, Finland, July 2001.

[174] F. Metelli. The perception of transparency. *Scientific American*, 230(4):90–98, 1974.

[175] J. Meyer. *Akustik und musikalische Aufführungspraxis*. Verlag das Musikinstrument, 1972.

[176] A. Michotte. The emotions regarded as functional connections. In M. L. Reymert, editor, *Feeling and emotions*, pages 114–126. McGraw-Hill, New York, 1950.

[177] S. K. Mitra. *Digital Signal Processing: A computer-Based Approach*. McGraw-Hill, New York, 1998.

[178] M. R. Moldover, J. B. Mehl, and M. Greenspan. Gas-filled spherical resonators: Theory and experiment. *J. of the Acoustical Society of America*, 79:253–272, 1986.

[179] B. C. J. Moore. *An introduction to the Psychology Of Hearing*. Academic Press, 4th edition, 1997.

[180] P. M. Morse. *Vibration and Sound*. American Institute of Physics for the Acoustical Society of America, New York, 1991. 1st ed. 1936, 2nd ed. 1948.

[181] P. M. Morse and K. U. Ingard. *Theoretical Acoustics*. McGraw-Hill, New York, 1968.

[182] A. Mulder. Getting a grip on alternate controllers. *Leonardo Music J.*, 6:33–40, 1996.

[183] E. D. Mynatt. Auditory presentation of graphical user interfaces. In G. Kramer, editor, *Auditory Display: Sonification, Audification and Auditory interfaces*, pages 533–555. Addison-Wesley, 1994.

[184] J. G. Neuhoff, G. Kramer, and J. Wayand. Sonification and the interaction of perceptual dimensions: Can the data get lost in the map? In *Proc. Int. Conf. on Auditory Display*, Atlanta, Georgia, USA, 2000.

[185] J. Nilsson and A. Thorstensson. Adaptability in frequency and amplitude of leg movements during human locomotion at different speeds. *Acta Physiol Scand*, 129:107–114, 1987.

[186] J. Nilsson and A. Thorstensson. Ground reaction forces at different speeds of human walking and running. *Acta Physiol Scand*, 136:217–227, 1989.

[187] E. G. Noik. Encoding presentation emphasis algorithms for graphs. In *Proc. Graph Drawing*, 1994.

[188] J. O'Brien, P. R. Cook, and G. Essl. Synthesizing sounds from physically based motion. In *Proc. ACM SIGGRAPH*, pages 529–536, Los Angeles, CA, 2001.

[189] H. Olsson, K. J. Åström, C. Canudas de Wit, M. Gäfwert, and P. Lischinsky. Friction models and friction compensation. *European J. of Control*, 4:176–195, 1998.

[190] C. E. Osgood. On the nature and measurement of meaning. *Psychological bulletin*, 49:197–237, 1952.

[191] C. Palombini. Pierre Schaeffer - from research into noises to experimental music. *Computer Music J.*, 17(3):14–19, 1993.

[192] A. Papoulis. *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, New York, NY, 1984. 2nd edition.

[193] A. P. Pentland. Fractal-based description of surfaces. In W. Richards, editor, *Natural Computation*, pages 279–298. MIT Press, Cambridge, MA, 1988.

[194] P. Q. Pfordresher and C. Palmer. Effects of delayed auditory feedback on timing of music performance. *Psychological Research*, 66(1):71–79, 2002.

[195] A. Pirhonen, S. A. Brewster, and C. Holguin. Gestural and audio metaphors as a means of control for mobile devices. In *Proc. ACM CHI*, pages 291–298, Minneapolis, MN, 2002. ACM Press, Addison-Wesley.

[196] Portishead. Dummy. CD, 1994: 828 553. London: Go! Discs.

[197] M. Puckette. Pure Data. In *Proc. Int. Computer Music Conference*, pages 269–272, San Francisco, 1996.

[198] Rane ttm 56, 2001. http://www.rane.com/djcat.html#mixersttm56.html.

[199] M. Rath, F. Avanzini, N. Bernardini, G. Borin, F. Fontana, L. Ottaviani, and D. Rocchesso. An introductory catalog of computer-synthesized contact sounds, in real-time. In *Proc. Colloquium of Musical Informatics*, Florence, Italy, May 2003.

[200] B. H. Repp. The sound of two hands clapping: an exploratory study. *J. of the Acoustical Society of America*, 81(4):1100–1109, April 1987.

[201] B. H. Repp. Acoustics, perception, and production of legato articulation on a computer-controlled grand piano. *J. of the Acoustical Society of America*, 102(3):1878–1890, 1997.

[202] S. Resnick. *Adventures in Stochastic Processes*. Birkhäuser Boston, 1992.

[203] J. C. Risset and D. L. Wessel. Exploration of timbre by analysis and synthesis. In D. Deutsch, editor, *The psychology of music*, pages 113–169. Academic Press, 2nd edition, 1999.

[204] D. Rocchesso. The ball within the box: a sound-processing metaphor. *Computer Music J.*, 19(4):47–57, Winter 1995.

[205] D. Rocchesso. Spatial effects. In U. Zölzer, editor, *DAFX: Digital Audio Effects*, pages 137–200. John Wiley & Sons, Chichester, UK, 2002.

[206] D. Rocchesso, R. Bresin, and M. Fernström. Sounding objects. *IEEE Multimedia*, 10(2), april 2003.

[207] D. Rocchesso and P. Dutilleux. Generalization of a 3D resonator model for the simulation of spherical enclosures. *Applied Signal Processing*, 1:15–26, 2001.

[208] D. Rocchesso and L. Ottaviani. Can one hear the volume of a shape? In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pages 115–118, NewPaltz, NY, October 2001. IEEE.

[209] V. Roussarie, S. McAdams, and A. Chaigne. Perceptual analysis of vibrating bars synthesized with a physical model. In *Proc.* 135^{th} *ASA Meeting*, New York, 1998.

[210] R. Saletti. A comparison between two methods to generate $1/f^{\gamma}$ noise. In *Proc. IEEE*, volume 74, pages 1595–1596, November 1986.

[211] L. Savioja, J. Backman, A. Järvinen, and T. Takala. Waveguide Mesh Method for Low-Frequency Simulation of Room Acoustics. *Proc. Int. Conf. on Acoustics*, pages 637–640, June 1995.

[212] G.P. Scavone, S. Lakatos, and C.R. Harbke. The Sonic Mapper: an interactive program for obtaining similarity ratings with auditory stimuli. In *Proc. Int. Conf. on Auditory Display*, Kyoto, Japan, 2002.

[213] P. Schaeffer. Etude aux chemins de fer, 1948.

[214] M. R. Schroeder. *Fractal, Chaos, Power Laws: Minutes from an Infinite Paradise*. W.H. Freeman & Company, New York, NY, 1991.

[215] Scratchdj. http://www.scratchdj.com, 2002. American turntablism site with forum, 2002.

[216] J. P. Sethna, K. A. Dahmen, and C. R. Myers. Crackling noise. *Nature*, (410):242–250, March 2001.

[217] B. Shneiderman. Direct manipulation: A step beyond programming languages. *IEEE Computer*, 16(8):57–69, 1983.

[218] B. Shneiderman. *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. Addison-Wesley, MA, USA, 2 edition, 1992.

[219] B. Shneiderman. Tree visualization with tree-maps: A 2-D space-filling approach. *ACM Trans. on Computer Graphics*, 11:92–99, 1992.

[220] Shure m44-7 cartridge & v15 vxmr cartridge. http://www.needlz.com/m44-7.asp, http://www.shure.com/v15vxmr.html.

[221] K. Sjölander and J. Beskow. Wavesurfer - an open source speech tool. http://www.speech.kth.se/wavesurfer/.

[222] M. Slaney and R. F. Lyon. On the importance of time – a temporal representation of sound. In M. Cooke, S. Beete, and M. Crawford, editors, *Visual Representations of Speech Signals*, pages 409–429. John Wiley & Sons, Chichester, UK, 1993. Available at http://www.slaney.org/malcolm/pubs.html.

[223] J. O. Smith. Physical modeling using digital waveguides. *Computer Music J.*, 16(4):74–91, 1992.

[224] R. Smith. *Online 1200*. Technics SL1200 Mk2 specifications.

[225] The Sounding Object web site, 2001. Containing software, demonstrations, and articles, http://www.soundobject.org.

[226] E. Somers. Abstract sound objects to expand the vocabulary of sound design for visual and theatrical media. In *Proc. Int. Conf. on Auditory Display*, Atlanta, Georgia, April 2000. Georgia Institute of Technology.

[227] R. D. Sorkin, D. E. Robinson, and B. G. Berg. A detection-theoretic method for the analysis of visual and auditory displays. In *Proc. Annual Meeting of the Human Factors Society*, volume 2, pages 1184–1188, 1987.

[228] R. Stallman. *Free Software, Free Society: Selected Essays of Richard M. Stallman*. The GNU Press, Boston, MA, 2002.

[229] J. Strikwerda. *Finite Difference Schemes and Partial Differential Equations*. Wadsworth & Brooks, Pacific Grove, CA, 1989.

[230] A. Stulov. Hysteretic model of the grand piano hammer felt. *J. of the Acoustical Society of America*, 97(4):2577–2585, Apr 1995.

[231] J. Swevers, F. Al-Bender, C. G. Ganseman, and T. Prajogo. An Integrated Friction Model Structure with Improved Presliding Behavior for Accurate Friction Compensation. *IEEE Trans. Autom. Control*, 45:675–686, 2000.

[232] S. Targett and M. Fernström. Audio games: Fun for all? all for fun? In *Proc. Int. Conf. on Auditory Display*, Boston, MA, July.

[233] M. Tirassa, A. Carassa, and G. Geminiani. A theoretical framework for the study of spatial cognition. In S. O'Nuallain, editor, *Spatial cognition. Fondations and Applications*. Benjamins, Amsterdam, 2000.

[234] T. Tolonen and H. Järveläinen. Perceptual study of decay parameters in plucked string synthesis. In *Proc. 109-th AES Convention*, Los Angeles, September 2000. Available from http://www.acoustics.hut.fi/publications/.

[235] B. Truax. *Acoustic communication.* Ablex, Norwood, NJ, 84.

[236] P. Valori. *Fenomenologia*. Sansoni, Firenze, 1967. Entry of the Encyclopedia of Philosophy.

[237] K. van den Doel, P. G. Kry, and D. K. Pai. Foleyautomatic: Physically-based sound effects for interactive simulation and animation. In *Proc. ACM SIGGRAPH*, August 2001.

[238] K. van den Doel and D. K. Pai. The sounds of physical shapes. *Presence*, 7(4):382–395, August 1998.

[239] N. J. Vanderveer. Acoustic information for event perception. In *Paper presented at the celebration in honor of Eleanor J. Gibson*, Cornell University, Ithaca, NY, 1979.

[240] N. J. Vanderveer. *Ecological Acoustics: Human perception of environmental sounds*. PhD thesis, 1979. Dissertation Abstracts International, 40, 4543B. (University Microfilms No. 80-04-002).

[241] Samurai series mixers. http://www.vestax.com/products/samurai.html.uk/, 2002.

[242] G. B. Vicario. On Wertheimer's principles of organization. *Gestalt Theory*, (20), 1998.

[243] G. B. Vicario. *Psicologia generale [General Psychology]*. Laterza, Roma, 2001.

[244] G. B. Vicario. Breaking of continuity in auditory field. In L. Albertazzi, editor, *Unfolding perceptual continua*. Benjamins, Amsterdam, The Netherlands, 2002.

[245] G. B. Vicario. La fenomenologia sperimentale esiste. *Teorie e modelli*, (7), 2002.

[246] G. B. Vicario. Temporal displacement. In R. Buccheri and M. Saniga, editors, *The nature of time: geometry, physics and perception*, Tatránska Lomnica, May 2003. NATO, Kluwer, Amsterdam, The Netherlands. In press.

[247] J. von Uexküll. *Streifzüge durch Umwelten der Tieren und Menschen*. Rowohlt, Reinbeck bei Hamburg, 1956.

[248] T. Waits. Mule variations. CD, 1999: 86547. Los Angeles, CA:Anti/Epitaph.

[249] S. Wake and T. Asahi. Sound retrieval with intuitive verbal expressions. In *Proc. Int. Conf. on Auditory Display*, Glasgow, Scotland, 1998.

[250] W. H. Warren, E. E. Kim, and R. Husney. The way the ball bounces: visual and auditory perception of elasticity and control of the bounce pass. *Perception*, 16:309–336, 1987.

[251] W. H. Warren and R. R. Verbrugge. Auditory perception of breaking and bouncing events: a case study in ecological acoustics. *J. of Experimental Psychology: Human Perception and Performance*, 10(5):704–712, 1984.

[252] C. A. Wert. Internal friction in solids. *J. of Applied Physics*, 60(6):1888–1895, 1986.

[253] R. Wildes and W. Richards. Recovering material properties from sound. In W. Richards, editor, *Natural Computation*, pages 356–363. MIT Press, Cambridge, MA, 1988.

[254] C. Williamson and B. Shneiderman. The dynamic homefinder: evaluating dynamic queries in a real estate information exploration system. In *Special Interest Group on Information Retrieval (SIGIR)*, 1992.

[255] F. Winberg and S.-O. Hellström. Qualitative aspects of auditory direct manipulation - a case study of the towers of hanoi. In *Proc. Int. Conf. on Auditory Display*, Espoo, Finland, July 2001.

[256] G. W. Wornell. Fractal signals. In V. K. Madisetti and D. B. Williams, editors, *The Digital Signal Processing Handbook*, chapter 73. CRC Press and IEEE Press, 1998.

[257] P. Zahorik. Assessing auditory distance perception using virtual acoustics. *J. of the Acoustical Society of America*, 111(4):1832–1846, 2002.

[258] P. Zahorik. Auditory display of sound source distance. In *Proc. Int. Conf. on Auditory Display*, Kyoto, Japan, July 2–5 2002.

[259] J. Zorn. Locus solus. CD Reissue, 1995: 7303. New York: Tzadik.