Università degli Studi di Verona
Dipartimento di Informatica

DOTTORATO DI RICERCA IN INFORMATICA
CICLO XVI

INTERACTIVE REALTIME SOUND MODELS FOR HUMAN–COMPUTER INTERACTION

A SOUND DESIGN CONCEPT AND APPLICATIONS

Coordinatore: Prof. Andrea Masini
Supervisore: Prof. Davide Rocchesso

Dottorando: Matthias Rath

"Seit ich hören kann, bin ich eingebunden in die Welt. Die Dinge reagieren akustisch auf mich. Sehen Sie diesen Löffel? Ich weiss genau, wie er aussieht, wie er sich anfühlt. Ich hatte ihn tausendmal in der Hand. Aber jetzt höre ich ihn, wenn er an die Tasse schlägt oder im Tee rührt. Er antwortet mir! Oder meine alte Jacke, die raschelt jetzt, wenn ich sie anziehe. Auch sie antwortet mir!

Die ganze Welt gibt mir jetzt Antwort.

Mein Laptop — jeder Buchstabe macht ein Geräusch. Das Geklapper hat mich anfangs so gestört, dass ich beim Schreiben keinen klaren Gedanken fassen konnte.

Wissen Sie — Hören ist eine Daseinsbestätigung für eine Person. [1]

Seit ich höre, begreife ich, dass früher [2] *die Selbstmordrate bei Späteretaubten zehnmal höher war als bei Späterblindeten:*

Der Ertaubte war von der Welt abgeschnitten."

Since I've been able to hear I've been integrated into the world. Things react acoustically to me. Do you see this spoon? I know exactly what it looks like, how it feels. I held it in my hand a thousand times. But now I hear it when it hits the cup or stirrs the tea. It is answering me! Or my old jacket, it now rustles when I put it on. It too is answering me!

The whole world now gives me response!

My laptop — every key makes a noise. In the beginning the clacking used to disturb me so much that I couldn't hold one clear thought when writing.

You know — Hearing is a confirmation of existence for a person. [3]

Since I've been able to hear, I can comprehend why the suicide rate among people that have become deaf was [4] *ten times higher than among people that have become blind:*

A deaf person was isolated from the world.

Maike Stein, born deaf, about her experiences of auditory perception by means of a cochlear implant (received at the age of 32 years, *Die Zeit* [63])

[1] Satz stammt ursprünglich von Aron Ronald Bodenheimer.
[2] Erklärung: bezieht sich auf 'vor der Entwicklung von Cochlear-Implantaten'.
[3] The initial author of this sentence is Aron Ronald Bodenheimer.
[4] explanation about context: "was" refers to 'before the availability of cochlear implants'.

"Dort am Klavier, lauschte ich ihr,
und wenn ihr Spiel begann, hielt ich den Atem an."

There at the piano, I would listen to her,
and when her playing began, I would hold my breath.

Rammstein [5]

"Bring the noise."(!)

Public Enemy [6]

[5] Rammstein: *Sehnsucht, Klavier* 1997
[6] Public Enemy: *It takes a nation of millions to hold us back* 1988

Acknowledgements

I wouldn't have been able to do this doctorate and especially to finish this thesis, without the presence, help and support of many people, some of which I'd like to mention here (in the hope not to get too sentimental thereby).

I'd like to thank my supervisor, Davide Rocchesso, for putting trust in me and hiring me for *SOb* [7], the European project that was so great to work for ("Party people! — Can you get phonky?! — Yeeaahhh!"), for all his obligingness, such as enabling me to conclude the intermediate diploma of my previous studies of electronic composition in Essen and giving me the chance to do the doctorate; finally for demonstrating that it is possible to procede doing research at a university while keeping a fresh, open mind and visible enthusiasm as the original motivation of one's work.

Warm greetings to the people at the SAP division of *ifb* [8] Köln! It was exceptionally pleasant to work with you — representing I just name my old–time *studying* (ehm...) colleague Jens Hielscher. The very fair employment at *ifb* made it possible for me to start studying again and deepen my fascination for electronic sound processing, thus to finally steer my life in the direction of this PhD.

Giovanni De Poli inspired me to come to Italy by notifying me of the possibility of participating in research projects, spontaneously after some question of mine at the *ICMC* 2000 in Berlin, and by forwarding the mails that finally led me to come to Verona. Thanks for his review of this thesis and for giving me time to finish it before starting to work in Padova.

Laura Ottaviani, one of my officemates (all of which were the most pleasant you could wish for) and fellow doctorands, guided me safely through the always astonishing world of a foreign bureaucracy when arriving. (Without her help I surely would never have gotten hold of my *codice fiscale*.) I have to thank her and Anna Maria Baltieri for patience and help with omnipresent administrative traps that I was about to fall into (due to chronically missing mutual respect and understanding, both linguistically and psychologically). Thanks for understanding and support with my troubles in the end phase of the doctorate also to our PhD coordinator Andrea Masini.

[7] "*SOb*" (note the setting) is the acronym of the European research project that I've been working for, "*The Sounding Object*" (nothing else — 'Son of...').

[8] *ifb* AG, http://www.ifbag.com/

Roberto Bresin initiated my 8-months visit to *KTH* (*Kungliga Tekniska Högskolan*/Royal Institute of Technology), Stockholm. To him and the very nice people at *TMH* (*Tal, Musik och Hörsel*/Department for Speech, Music and Hearing), in particular Kjetil Falkenberg Hansen, Sofia Dahl and Sten Ternström, I'm very grateful for my stay in Sweden as a rich experience, personally, culturally and professionally.

If the *C*-code that I wrote during the docorate and for the *Sounding Object* not only compiles and runs somehow, but is also manageable and fulfills minimum standards of structuring and readability, it is due to the lessons on the praxis and ethics of programming by Nicola Bernardini ("Porca miseria!..."). Btw.: I think we still have to conclude our discourse on the hedonist value of Claude Young's mix LP and this "techno music" in general...) and Gianpaolo "Defeater of the Pentium Bug" Borin (Thanks for forcing me to add some comments to that code!).

For motivating and enlightening remarks on my work I thank Joël Bensoam and Nicolas Misdariis from *IRCAM* (Institute de Recherche et Coordination Acoustique/Musique, Paris, France), standing also for stimulating encounters in the past three years with many other people, all of whose names I can not list here.

In his review of this thesis (although delivered to him piecewise in a 0.01 beta gamma version), Christian Müller-Tomfelde invested exceptional energy and competence to give me feedback about the main thoughts of the work as well as many details (Of course you're right: it's a red *cross*, not a *point*...). I'm very thankful for his deep and helpful comments.

I thank Maike Stein for her kind (and quick — answering to my mail during her summer holidays...) permission to cite from her interview with *Die Zeit* (page iii) and for her explanation about A. R. Bodenheimer as the original source of one of the cited phrases (see footnote on page iii).

Thanks to Umberto and Paola Castellani and friends, Linda Brodo, Marco Cristani and Manuele and Giovanna Bicego for welcoming so warmly a confused little *Kraut* in Verona. Thanks to Arnaud Fleury for helping me a lot by finding an appartment back then, and for his *fantastic* introduction into the calligraphic tradition of the French-speaking heterosexual Calvinist community of Malaysia. It will to a large part be the merit of Fausto Spoto's never-ending struggles to integrate isolated life forms found in Verona, if higher biological evolution can be found also amongst the students of the Department of Computer science (Jesus is with your fight for a bigger parking place!...).

Federico Fontana already owns an everlasting rank in the history of science for proposing, founding, organizing and hosting the now classic *CCCC* (*Christmas Conference on Chinese Cooking* [9]). Personally, I'm further on very thankful for help and exchange, whether concerning engineering knowledge, problems of my Italian (thanks a.o. for translating some of the "Sommario"), general scientific discussions or administrative advice, and for establishing my connection

[9] Venezia, December 2001; Venezia, May 2003; Köln, Karneval 202?; Bejing, year of the 12 monkeys...

to the Department of Information Engineering (*DEI*), Padova. Thanks for help and explanations to Amalia De Götzen and for importing some additional avant-gard spirit. I always truly enjoy(ed) cooperating with Federico "II" Avanzini, another *CCCC* founding member, on daily research work, presentations and papers and rather informal occasions.

I'm grateful to Nathalie Vermeire for her native speaker's advice with my biggest doubts concerning English formulations in some exposed parts of this thesis. (The surely many remaining linguistic oddities are exclusively my work.)

Deepest respect to Bruno Giordano for proving that true passion for phenomena of the mind and world (don't worry: no details here...) remains a stronger force than the power of opportunism and the desire for quick superficial success (yeah); and for showing in his work the humbleness and *Gründlichkeit* (thoroughness) that are so rare among the young people of today (sigh...). My gratefulness also for 200 GB of matlab scripts and mp3s as well as ca. 117 pool games all of which helped me to overcome the frustrations of a lost PhD student's existence.

Thanks to Samir Genaim for his constant spontaneous helpfulness, in too many situations to list here, and for beautiful Arabian music and food.

A big thankful embrace to Linmi Tao who repeatedly kept me alive through his unceasing production of *delicious* chinese food, helped me out in any occasion and gives a notorious European a demonstration of *Tao*ist wisdom. He also receives the *CCCC* medal of best (and only) presentation at *CCCC* 2001 and 2003.

Thanks to Giuditta Franco for filling the fog that tendentially surrounds the Department of Computer Science in Borgo Roma, with bright Calabrian sunshine, for infinite patience with a stupid German trying to speak Italian (Grrahtsie!! Tschau, tschaou...), for demonstrating that shining intelligence and wit (which will never be substitutable by ambition and discipline...) can combine with equally impressing personal warmth and charm and for proving the absolute power of female wisdom.

Whenever I seriously needed to tank some good old *Rheinland*-spirit, the 5th floor apartment of my long–time musician (well: drummer...;-) friend Marc Bierbaums was the safe refuge to go for. Cheers to him for always welcoming me!

The main factor that led me to do those strange things that I do is surely my family. I thank my aunt Anneliese who always supported my struggles and inspired my love for art as well as nature and science in an immeasurable way (and whose level of daily creativity is a high benchmark), my sister Katharina who cares for me more than I deserve, and most of all my parents who allured me to look beyond the surface and in all directions, and without whose logistic and material support my recent nomadic lifestyle would have ended in a fiasko (much earlier).

Contents

Introduction

If one considers the omnipresent importance and impressive capabilities of human auditory perception as one of the two main human senses together with visual perception, the rather peripheral and primitive role of auditory display in most human–computer interfaces today is even more noticeable. Although sound has been recognized as an effective channel for the transmission of information in human–computer interaction (see e.g. [17][15][18]), its use is still mostly restricted to short, fixed and static, usually prerecorded, signals; most familiar are such of a warning or notification-character. Even where sounds of longer temporal duration are used, e.g. acoustic "atmospheres" in computer games, they are generally not reactive, but almost solely played-back *sound samples*. This situation, that leaves one big part of human resources of perception, communication and interaction unused, and that surprises also on the background of the long and broad musical adoption of electronic sound production [10], is more and more recognized as poor, restrictive and unnatural. (One may make aware, that in the "real world" we virtually always hear, also when we "don't look at" or "don't pay attention", while most current computer systems force us to constantly stare at displays.) Increasing efforts have been spent lately on the development of new, enhanced ways of auditory display. Reasons for this growing interest lie also in recent trends in the science and practice of informatics: an adequate sonic component is seen to be of high significance for convincing environments of virtual or augmented reality [10]; and wherever computing power is to be further integrated and "mobilized", and computers as distinct units disappear, e.g. under the premises of *ubiquitous*, *pervasive* or *wearable* computing, where the facilities of graphical displays are restricted, the auditory channel is understood to be the first–choice alternative and support. Much work has been spent in the area of sound *spatialization* and *localization* [7], whereas audio *sources* have received far less attention so far.

Maybe one of the main factors connected to the "underdeveloped" situation of auditory display is the traditional focus of psychoacoustic research on abstract properties of sound and acoustic signals. While long dedicated research has resulted in well-established connections between specific signal-theoretic properties, such as frequency and amplitude, and (conventional) musical terms, such

[10] Electronic sound has probably a much longer and prominent tradition in music than the use of computers in graphical arts...

as pitch and loudness, the perception of "non-musical", everyday sounds has been examined much less. Simultaneously, older methods of sound synthesis, such as subtractive, or FM synthesis, are based on, and controlled in terms of, parameters of signal theory, and are in their resulting sonic character generally quite restricted and distinct from the sounds of our usual surroundings. The use of *sample-based* sound, i.e. the common playback of (possibly modified) prerecorded sound files, that forms the current standard in computer interfaces and tools, can be seen as the first reaction on the, already mentioned, restrictions of traditional techniques of sound synthesis, with their signal-theoretic control parameters. *Sampling* however is unsatisfactory in many respects, for its static, not reactive nor dynamic sound character. But all these previous obstacles for the opening of new paths of enhanced auditory displays have started to be, and are more and more being, dissolved through recent developments in both, psychoacoustics and sound generation.

Foundations for a sound design concept

In the field of psychoacoustics, the *ecological school* points out that human auditory perception in our everyday surroundings is of different nature than the sensation of abstract, musical or signal-based attributes. It has been noted [74] [29] that especially non–expert listeners, i.e. average listeners with low preparatory training in music and acoustics, tend to describe sounds they hear in terms of possible *sound sources* and their attributes, and only exceptionally (e.g. when confronted with unusual sounds that are very hard to classify) refer to abstract sound properties such as pitch, brightness or loudness. This observation is reflected in the introduced terms of *everyday listening* as opposed to *musical listening* [29]. The human capability and tendency to extract ecological attributes from sound has been subject of an increasing number of psychoacoustic studies. Uncovered and examined has been the auditory perception of *transformational invariants*, i.e. attributes of sound-emitting processes in our surrounding, such as velocities, as well as *structural invariants* [30], i.e. attributes of involved objects, such as material, size or shape. Such works lay the basis for respective efforts of *ecologically expressive* sound synthesis. These do not necessarily have to result in imitations of exemplars of "real" sounds from everyday surroundings. To be intuitively informative and expressive in an unambiguous, clear or stressed way, it may generally be desirable to isolate or (over-)exaggerate in auditory display, certain important ecologic attributes of a complex familiar scenario, on the cost of others considered of minor interest. The term of "*cartoonification*" is used to refer to such a clearer, focused auditory *ecological expression*, in allusion to graphical cartoon icons that can, while being clearly recognized as artificial, represent "real" objects or certain of their attributes, often more unmistakeably than photorealistic pictures.

Achievements of the *ecological approach* in psychoacoustics have been reflected in some according results of auditory display, that demonstrate the applicability and potential of the innovations described above (see e.g. [29], [30], [31], [51], [71], [73]). However, there is still much free space for further re-

spective efforts of sound generation; and the formulation and exploration of a more general, systematic technical approach towards the practicable realization and exploitation of the ideas of *ecological auditory expression* and *cartoonification* is a worthwhile goal. In particular, a deeper, systematic connection of various existing, including newer, techniques of sound synthesis and the psychoacoustic approach mentioned above, considering also aspects of usability and implementation, has not been established. This may also reflect typical roles of auditory and visual expression. Sound is generally recognized in its enormous relevance as the medium of language and music. But, while every child knows how to draw "smileys" or other cartoon icons, basic orientation or concepts of how to approach *ecologically expressive*, efficient sound design are still needed. Of high interest from the *ecological* standpoint is a rather recent tendency in sound generation, known under the term of "*physical modeling*" and based on physical–mathematical descriptions of (e.g.) mechanical sound emitting systems, rather than properties of signals (to be generated). Physics-based synthesis algorithms are in their control parameters naturally connected to *ecological attributes*. The largest part of works in the field however, is concerned with the possibly realistic simulation of single, unique physical systems, mainly musical instruments. Resulting implementations are tendentially too complex in control and computation for the use as part of a human–computer interface [11], and usually highly specialized and rather inflexible in their sonic potential. Traditional musical instruments in fact, can be seen as efforts to "hide away" their ecological character, addressing *musical listening* (surprisingly...) rather than *everyday listening*. A deeper, dedicated link, joining the experience of physics-based sound synthesis and the insights of *ecological psychoacoustics* is only recently being developed [12]. In particular, up until recently, the notion of *cartoonification* has not deliberately and consequently been introduced into physics-based "*sound modeling*".

General points of this thesis

The work presented in the following forms a path to overcome or improve the unfortunate and unsatisfying current situation of auditory display. Tools are provided and a sound design concept is set up and reproved, to enrich human–computer interaction through an enhanced, new use of the auditory channel, adequate to the indispensable, uninterrupted human perception of acoustic information in "natural" surroundings. For auditory display to be intuitive, in the sense of being spontaneously understandable and steering a user's (re)actions without previous explanation or training, the aim is *ecological expression*, as opposed to abstract sound signals. The central idea of *sound modeling* can be

[11] Of course, the audio channel of the interface of a system can not consume the same amount of computational resources that a stand–alone dedicated device, e.g. an electronic musical instrument can rely on.

[12] ... e.g. in the course of the European research project "*The Sounding Object* (SOb)" [67] that the author of this thesis has been working for, and which has strongly influenced and inspired the work presented here.

seen as the auditory pendant to the creation of graphical representations (such as icons or cartoons) of known, familiar objects or scenarios. Further on, the *sound models* presented here incorporate a dynamic (complex) sonic behavior rather than (collections of) fixed, isolated signals. They overcome the restrictions of sample-based sound in its static, repetitive nature, and provide reactive sonic feedback that can instantaneously express and reflect ongoing processes, e.g. connected to streams of user input. The principle of *cartoonification* is extensively applied to combine clear, focused expression with affordable realtime implementation. To stick with the analogy of graphical display, one may compare graphical icons or cartoons, e.g. on a computer desktop or traffic signs, that are both, easier to draw (cheaper in "implementation") and clearer to comprehend in their meaning, than photorealistic pictures. Chapter 1 deals with this background, scopes and terminology in detail.

The described acoustic qualities are achieved by applying state–of–the–art techniques of sound synthesis, namely the use of physics-based models. This also provides acoustic results whose perceptual dimensions are not (yet) covered by signal-theory. However, abstractions are searched and derived where useful for the sake of flexibility, economy of computation and implementation and clearness of expression. In the process, experiences and strengths of conventional techniques of sound synthesis are not ignored, but instead exploited as well, resulting in an *hybrid architecture* that combines *physics-based* and also *signal-based* techniques in perception-oriented structures. Details of the concept are described in chapter 2. At all stages, human perception, understanding, action and use are the ultimate gauge to be followed, which justifies to use the term of sound "design" (rather than simply "synthesis" or "production").

As a consequence of their dynamical behavior and reactivity in realtime, the *sound models* can be naturally combined and synchronized with other perceptual modes, such as graphical display or gestural user input; some examples are presented in chapter 3. The solid embedding into clear, possibly familiar, overall metaphors for the interaction with, or control of, a system, can further consolidate intuitive understandability. This principle is exemplified in one of the multi-modal example devices (of chapter 3), the "*Ballancer*", an interactive tangible–audio-visual "game". Evaluation experiments described in chapter 4 show the suitability and success of the concept and development work of chapter 2 at the example of the "rolling" model and the *Ballancer*. These tests also prove and measure the improvement of user-performance through the exploitation of continuous informative and reactive sonic feedback, as present in our actions in everyday situations (and missing in current human–computer environments). The chapter on evaluation may claim uniqueness, in that respective results have not been demonstrated before, neither in general in this clarity nor in any concrete application as the one introduced here.

Structure and main achievements

- The motivation and bases in psychoacoustics as sketched above are displayed in chapter 1. On this basis, the general scopes of the following

work are explained, as well as the use or meaning of central terms such as "sound modeling" or "cartoonification".

- Chapter 2 contains the *sound modeling* work, starting from a general layout of the underlying concept in its main points, i.e. scopes and technical approach (section 2.1).
  - Section 2.2 gives an overview of the concrete application of the general concept on the major class of everyday sound-emitting processes, impact-based contact scenarios of solid objects.

  The hybrid, hierarchical architecture that is one part of my sound design approach is reflected in the two "technical" sections 2.3 and 2.4.

  - Section 2.3 contains the development and implementation of the "straight" physics-based, low–level model of solid objects in impact-interaction. The main new achievements here are the integration of *modal synthesis* with a dynamic physics-based description of impact, and the *modular* realtime implementation (section 2.3.3).
  - The realization of more complex scenarios in higher–level structures that control and make use of the developed underlying audio kernel is presented in section 2.4. These higher–level models are new results in the field of sound synthesis. In particular, approaches or a results in the realtime modeling of "breaking" (section 2.4.2) did not exist so far, and the model of "rolling" (section 2.4.3) allows to reach a degree of plausibility, ecological expression [13] and detailed realism not reached in previous works of synthesis.

- Chapter 3 contains some examples of the integration of sound models presented in chapter 2 with multi–modal interfaces.
  - Most important is here the *Ballancer* (section 3.3), an interactive tangible–audio-visual "game" of balancing a virtual ball on a track. The *Ballancer* is highly relevant for the thesis as a whole, in fact more than an example, since it is also used in the largest part of the evaluation experiments that are reported in chapter 4.

- In the last chapter (4) the suitability and success of the sound design work in reaching the initial scopes are demonstrated and evaluated, at the example of the most complex sound model, that of rolling, and including multi–modal interaction (through the *Ballancer*).
  - Section 4.2 presents the first part of the tests addressing the potential of the sound model to clearly represent a familiar scenario (rolling interaction) in itself and, as a result, to steer a user's understanding of, and interaction with a system, without additional explanation or training.

[13] The evaluation of the model in chapter 4 concretizes and justifies this characterization.

– The second part of the evaluation tests, reported in section 4.3, proves the continuous transmission of information through the sound of the model and its intuitive (i.e. spontaneous, unprepared, untrained) perception by users and exploitation in performance improvement. The experiment and analysis developed here are unique in that they allow to detect and expose unconscious mechanisms of auditory perception and interaction through detailed measurements of control movements. Bias through conscious reaction and reflection that would result from direct questions (as in previous literature) is minimized in this "indirect" strategy. The content and results of this section (4.3) indeed have a significance for psychoacoustic research that goes beyond the closer scope of approving the success of the sound modeling work in this thesis. The direct gestural exploitation of continuous sonic feedback has never been proven before and may thus form a basis for fundamentally new principles in auditory display, in the sense outlined in chapter 1 (section 1.1).

Remarks

The work presented in this thesis touches a rather wide range of fields of research. While the central point of focus lies in the provision and utilization of new principles and techniques of auditory display for human–computer interaction, knowledge, impulses and activities in the areas of psychoacoustics, sound synthesis, the modeling of physical systems, realtime-programming, sound and interface design and psychophysical evaluation are essential necessities to reach (and even understand and formulate) the scopes of this work. Vice versa, the achievements and various intermediate steps during the course of the project presented here can probably contribute and be of value and relevance for several of the mentioned and related fields of research. This should however not obscure the overall direction and progression of this thesis. All parts of the work, that root in or lead into subsidiary terrain, have to be seen as essential building blocks in the final constructions of auditory display for human–computer interfaces. The ultimate scope, that will be displayed and explained in depth (chapter 1), must always serve as the orientation mark, and especially sections of rather technical character (chapter 2) must be seen as significant, with the overall framework in mind, and not be misunderstood as ballast or digressions. On the other hand it must be understood that some developed approaches might surely be deepened and completed in the sense of specialized fields of research, which can sometimes not be done here because that would lead away from, and not contribute to, the direct point of interest.

Finally a remark has to be made concerning the relation of the concrete practical developments presented in the following and the wider concept, sketched above and elaborated in the course of the text. Of course I do not claim to have, once and for all, exhaustively realized (and evaluated) in all its possibilities a sound design approach based on *ecological expression*, *cartoonification* and the integration of physics-based and signal-based sound synthesis. The con-

crete developments in this thesis deal with contact sounds of solid objects, as the perhaps most important class of sound-emitting processes in our common surroundings. I focus on scenarios based on micro-events of impact-interaction, "hitting", "dropping", "bouncing", "breaking"... (chapter 2), out of which I further concentrate in application and evaluation on the model of "rolling" as particularly rich in its potential to transmit (ecologic) information. Contact sounds based on friction have been the subject of closely connected research [59]. I do not claim completeness in any respect; sounds of origins of gas e.g. form one of the related fields not touched in this work (see e.g. [22]) and also the objects covered here could be approached in various ways (as a consequence of perception-orientation, *cartoonification* and abstraction in addition to varying forms of physical descriptions). But the developed sound models are to be seen as carefully chosen instances to explain and substantiate a general approach of sound design. A pure theoretical concept would be worthless without approval in concrete realizations, just as an arbitrary collection of sound generation algorithms without a common higher concept, a solid structural basis, would be of minor value.

Chapter 1

Background and Scopes

1.1 Psychoacoustic bases

The sound models [1] developed in chapter 2 in many aspects connect to, continue and build upon a pioneering work of William W. Gaver, "Everyday listening and auditory icons" [29]. Psychology however is not the direct field here and sound in human–computer interaction is not a final application but the central and final scope of interest. As a consequence, I focus much more on technical questions and potentials of sound generation and practical implementation and look at psychological aspects only as deeply as necessary or helpful for this practical scope. Human perception is of interest here not so much as a phenomenon by itself but in its function to supply us with information. In particular I stress the connection between an adopted "*mode*" of perception, the potential information to be perceived and its potential to steer and enable actions and reactions, because this is the setting at hand: human–computer interaction. To give a clearer idea of this last point I start off below with two examples that may initially appear far-fetched; the real concrete relevance of the connection "perceptual *mode* – conveyed information – enabled/provoked interaction" will become clear through the results of the evaluation experiment in chapter 4.

I make use of, and thus shortly sketch or at parts cite, some main thoughts, ideas and terms introduced or described by Gaver. I try however to avoid getting deeper into questions of psychological theory and to use terms such as "information", "perception" or "representation" in a neutral manner with respect to different and opposing psychological standpoints, such as *ecological* versus *cognitive*.

What can psychoacoustics tell us?

The central motivation and goal of the thesis is to contribute to a deeper, more effective exploitation of the sonic channel in human–computer interaction. Very

[1] This term will be precisely defined in section 1.2.

generally, I am concerned with the

question 1 *"How can information in a computer environment be conveyed to a user through the acoustic channel?"*

At this point I use the term "information" in a possibly wide sense, standing for basically anything that may be of interest for, or somehow influence the behavior of a human user of a system (being it through emotional reaction, rational understanding or...). It may appear natural to approach the question 1 by first specifying more concretely the information or content that is to be transported, in other words to start by asking back and seeking a thorough answer to

question 2 *"What information is to be displayed?"*

On the other hand, knowledge about the potential of human auditory perception is necessary to have an a–priori idea of which kinds of information should and can be transferred through an auditory interface (rather than via alternative channels of perception, e.g. the visual): there is a close link between the employed perceptual channel and the nature of the transferred content or knowledge. I believe that this remark is all but negligible. To give an example, think about the photo of a human face (on a computer screen) that may enable us to identify an (otherwise unknown) person (without further help, such as his/her name...) with high security, e.g. within a large group of people. It may be impossible to reach the same performance of identification with a verbal description of the picture (black hair, brown eyes...); no matter how exactly the photo is described (verbally), there may be several persons that share the same formal characteristics, but still can be distinguished visually (from a photo of sufficient quality). In this case, there is clearly some information contained in the photo, "the visual identity" of the person, that can be perceived visually by a viewer, but can in no (known) form be perceived through the auditory channel: the information contained in the picture may be *encoded* acoustically without loss, e.g. the file of a digital photo can be *transferred* through a modem; but listening to and identifying such an acoustic representation will never enable the listener to *identify the person* of the photo — neither by looking directly nor with the help of, lets say, a digital camera connected to a modem. [2] Part of the information in the picture can not be *perceived* auditorilly. Just vice versa, and perhaps more striking, we can precisely distinct the voices of friends, but try to communicate the identity of a friend's voice to an outsider without using the auditory channel (e.g.) by playing back an audio recording of your friend!? You may write an exhaustive description (of accent, voice range...) or print measured waveforms or spectrograms, but can you produce a graphic or picture that will enable a stranger to identify your friend when hearing his voice, as you can (e.g. on the phone)? The result of this *Gedankenexperiment* ("theoretical experiment"), the coercive connection of content and its perception, is not simply a question of resolution: with some training, an expert listener may

[2] This is what common experience tells; of course I have no formal or experimental proof for this claim, and in fact I will come back to the point, seen in a slightly different light below.

be able to recognize an acoustic representation of the photo, e.g. the according modem signal, with very high security as well; but doing so, he would in no way learn to identify the depicted person. As well, you may memorize a high resolution waveform display of the vowels and consonants of a human voice, which will never enable you to recognize its owner when hearing him. But what then is the reason that the visual and auditory channel of perception can not easily be exchanged in the examples, that certain information is decoded only meaningfully visually or only by auditory perception? What is the origin of the problem, if it is not simply one of a picture being "too big to be heard" or a human voice being "too complex to be seen"? The answer obviously lies in fundamental differences of processing of information by the visual or auditory channel of human perception. Human visual perception is made for seeing faces not voices, as voices are to be heard not seen, by the nature of our perceptual system, not simply because they "reach the ear and not the eye". The latter, purely physical division, in fact can be overcome today through technical apparatus (such as a modem), yet not so the fundamental differences between vision and audition: we can make a voice *visible* (e.g. a waveform display) but not "*seeable*". One perceptual channel structures incoming information into units as "faces", "heads" or "eyes" while the other one uses structures such as "voices", "cries" or "rhythms". Maybe, if we had a deeper understanding of the processes and structures of visual and auditory perception, we might indeed construct realtime-converters that would enable us to "hear faces" and "see voices" in the above sense, comparable to braille printing for visually impaired that allow to read text through the tactile channel to the same extend as we usually do visually.(?) In any case, the existing verbal (or other formal) representations of visual and auditory percepts (above through terms as hair color or voice range) are obviously not strong enough to achieve such exchangeability.

Having looked in the last, somewhat rambling paragraph, at the linkage of perceptual channels and perceived information, I argue that also within one perceptual channel, namely the auditory one, that is at interest here, different "subchannels", mechanisms or "modes" may be present, that allow the perception of different incomparable or unexchangeable qualities, and that are connected to different objects and attributes. As examples, music and (spoken) language might be considered such subchannels, even if surely not as clearly separable as vision and audition: we all share the experience, that the perception of a piece of music, say a symphony, can not be satisfyingly communicated in words. There are traditional representations, such as scores with marks of instrumentation, that may allow to widely "reconstruct" musical pieces (e.g. by a performing orchestra), and might be read off. But reading a musical score [3] will usually never replace the experience of hearing music [4], just as a printable

[3] . . . or more exactly in this example: listening to somebody who verbally reads off or describes, a musical score,

[4] To be exact, many listeners of music will probably agree with this observation, while there exists also a somewhat provocative, contrary viewpoint ([48], chapters XXII or XVII). At this point however, I consider the latter idea as rather exceptional, closely linked to a certain cultural background of western music tradition.

waveform will not replace a heard voice, only encode it. From the considerations given so far, it should be understood, that I do not look at the central question 1 stated at the beginning as a pure engineering task. Psychological knowledge about auditory perception is not only needed to place the design of auditory display on a more solid basis than "trial and error", it is also necessary to guide and specify our a–priori expectations about the potentials and goals of auditory display. Starting point here is thus a look at results and ideas of psychoacoustic research rather than concrete applicational specifications [5], i.e. possibly detailed answers to question 2. Viewing impulses from psychoacoustics and the general demands in human–computer interaction simultaneously, the aim is to construct new tools and interfaces that can enhance the capacities of auditory display not only gradually: through the exploitation of previously unused mechanisms of auditory perception, it can be expected to transmit to the user of a system, information of qualitatively new types. In fact, in the final chapter (4) of evaluation, the auditory expression and perception of the momentary velocity of a virtual ball is demonstrated, that leads to significant performance improvements of test subjects in an interaction task. It is not known, how the same effect on performance, i.e. the same information flow could be established by using conventional approaches of auditory display or through other perceptual channels or modes (e.g. through vision or speech). The psychoacoustic approach that the work leading to the mentioned result (in fact all the work described in the following chapters) is based on, is the *ecological* one. [6] I try to give a minimal sketch of this psychoacoustic background (psychology is not directly the field of this thesis, although some results are surely of interest for psychologists...) and to state the main ideas, terms and works of relevance here.

Everyday listening and acoustic signals

From an ecological viewpoint, auditory perception (as the other perceptual channels), serves the function of delivering information about our surroundings, besides (and maybe first of all...) its obvious relevance through spoken language and music. This notion is not surprising on the background of biological evolution, e.g. when we assume that our biological ancestors already had ear–like organs long before any forms of speech and music existed, or that children already pay attention to sounds before they start to speak (which should be out of doubt). Surprising is rather the fact that we are generally not aware of the importance of auditory perception as a channel for ecological information: the difficulties in common surroundings connected to visual disablement seem obvious to everybody, while much less attention is directed towards the importance

[5] This latter direction is probably what mostly comes to mind when confronted with the term "*sonification*", the acoustic representation of data.

[6] The *ecological* approach in psychoacoustics appears to be distinguished from its *cognitive* counterpart in various aspects that I do not discuss here. Some of the following arguments might be classified by a psychologist as rather typical for a *cognitive* standpoint. I use the term "ecological" in its more direct sense as looking at listener and sound in their environment.

of the perception of environmental sounds. Reasons for this may lie in the omnipresence of sound and hearing — we can instantaneously verify the importance of visual information simply by closing our eyes or turning off the light, while we have no "earlids" — or the fact that music and speech necessarily come to mind when thinking about acoustic sensations (graphical arts are probably of less importance for most people and written language is derived from its spoken predecessor). Another reason may be that the "modern world" is dominated by visually transmitted information, which in turn reflects the lower attention towards sound as a source of information (besides speech and music), invested in society and by psychologists in particular. In fact, traditionally, psychological work on audition is mainly concerned with musical phenomena such as pitch or consonance or the perception of different (conventional) musical instruments. I believe that one factor responsible for this traditional focus (and the one-sided use of the auditory channel) is not so much of general cultural, but rather of technical nature.

The psychological "study of perception is to a large part one of the mapping of energy in the world and energy — and experience — in a perceiver" (Bill Gaver [29]). Since variations of air pressure at the eardrums are the necessary cause of auditory percepts [7], psychoacoustic research is interested in mappings of attributes of **1**. sound sources (i.e. of sound emitting processes), **2**. of the acoustic signal (i.e. the signal of time-varying air pressure) and **3**. of human experience. The maybe most groundbreaking and significant step) in this respect was made by Helmholtz [75] [76] when connecting the "sensations of tone" and parameters of the representation of periodic signals as series of sinusoids after Fourier. The fascination of Helmholtz's results at initial contact is easy to share, even more on the background of previous, older knowledge about auditory perception (e.g. Pythagoras' relations of the lengths of vibrating strings and prominent musical intervals). Since Helmholtz's original work, his idea of predicting sound experiences by (or connecting them to) parameters of Fourier–related, *spectral* representations of acoustic signals, has been extensively carried on and used for sound manipulation and generation, with some impressive success. (It is basically impossible today to find or lead a discourse concerning acoustic signals and sound without at least the mentioning of spectral signal attributes.) On the other hand, it appears that Fourier–, or related, e.g. wavelet-based, techniques have at times been seen as the absolute, ultimate and omnipotent principle to explain auditory perception. An example is the still widespread idea of the human ear being "phase-deaf", that goes back to one of Helmholtz' original claims and that is connected to a view of acoustic processing in the outer and middle ear known as "*place theory*". The transfer of the signal of air pressure at the ear to movement of the cochlea is here seen a form of windowed Fourier transform and the following neuronal or cognitive stages of auditory perception are assumed to process only the information of

[7] I here ignore such phenomena as the sensation of acoustic vibration at other parts of the human body (low frequencies) and auditory percepts without mechanical movement such as tinnitus tones as subsidiary.

the maximal or average activity along the length of the cochlea [8] and not the exact temporal behavior; phase shifts of spectral components would thus not be perceivable, for reasons located already "before" the inner ear. The latter belief has repeatedly been disproven [64] [53] [54], and while cochlear processing indeed may be approximated by a wavelet transform [20], the involved temporal windows are short (in comparison to the periods of audible frequencies) and phase information about the movement of the cochlear at the different places is available for further stages of processing and obviously also relevant (for the resulting sound experience, at least for part of the cochlear). In consequence, Fourier- or wavelet-based spectra may represent the rough "preprocessing" in the middle ear but not auditory perception in general. They seem to be suitable for the explanation and manipulation or control of rather rough auditory attributes such as "brightness" and for (parts of) signals of specific characteristics such as signals with only or mainly pseudo-periodic components. The auditory attribute of pitch (with its "derivations" such as harmonicity) in fact appears exceptional as the clearest [9] auditory sensation that is strictly related to one rather straightforward attribute of acoustic signals, periodicity; perceived pitch is quite well predictable from suitable Fourier spectra, although the process of its perception is not solely based on a mechanical one in the outer and middle ear []. The auditory perception of most environmental processes and attributes in contrast seems quite hard to explain in terms of parameters of Fourier- and wavelet-transforms. E.g., it is doubtable if any spectral representation (whether Fourier– or wavelet–like) of short transient parts of contact sounds can be very helpful in predicting/explaining their auditory perception, more than the direct temporal representation of such signals. Moreover, for many phenomena of perception of everyday sounds it can be doubted if a satisfactory reduction or explanation in terms of parameters of mathematical transforms comparable to those known under the name of Fourier will ever be found. It might simply be a fact to be accepted that the perceptual processes leading to the identification of many everyday scenarios involve memory and subprocesses of various, e.g. statistical, nature [10] in complex connection, and can not be satisfactorily modeled, not even approximated by "homogeneous" mathematical operations in the conventional sense. As an example, a sufficiently reliable model of the recognition of sound as coming from a source involving fluids might necessarily consist of such a complex algorithm that it would be not more enlightening and useful (in the practice of sound design or in further psychological research) than judging by personal listening or dedicated statistical tests. This hypothesis is not meant to devaluate the powerful developed tools of sound processing, but to encourage sound design and psychoacoustic research not to rely entirely on Fourier–related techniques. For sound design I believe that making use of long–term direct, *intuitive*, *personal* experience in sound synthesis and listening(!) is indeed legitimate, moreover demanded — surely not as the only (or main)

[8] . . . in the different "places" on the cochlea. . .

[9] I consider e.g. brightness as less unambiguous in its definition and assessment.

[10] A psychologist would probably classify this argumentation as typical for a *cognitive* rather than an *ecological* approach.

basis, but as one in addition to, and possibly beyond the range of, traditional psychoacoustic knowledge. [11] The tendential "absolution" of Fourier–related parameters is reflected by the occasional use of the summary term "sound attributes" or also "attributes of the sound itself". From the viewpoint given just before, perceptual attributes, of whatever kind (and defined by whatever, e.g. statistical, mean), e.g. the *auditorily perceived material* [12] of an object, might be called "attributes of the sound" with the same right as Fourier–related parameters, since they are obviously derived from the acoustic signal [13], admittedly possibly in a process much harder to formalize mathematically; the latter ones might in this sense better be called "analytical signal attributes".

Without getting further into details of terms and definitions, fact is that human auditory perception has capabilities and a tendency to detect and assess from heard sounds their sources in our surrounding (as discussed in the following) and that these processes of auditory perception of everyday sound sources are often not satisfactorily described in terms of the Fourier–related tools of classic psychoacoustics. Vanderveer [74] has first observed that listeners tend to describe sounds they are confronted with in terms of attributes of sound sources and to rely on abstract attributes clearly related to the classical parameters of signal acoustics, such as pitch/frequency or loudness/intensity, only when they can not easily relate a known source. Accounting for (and stressing) the differences of the auditory perception of environmental sounds and of the acoustic parameters as mentioned, Gaver has introduced [29] the terms "*everyday listening*" and "*musical listening*". The latter is defined as the "perception of structures created by patterning attributes of sound itself" [14] or concerning attention to the "*proximal stimuli*" while "everyday listening involves attending to the distal stimuli", i.e. the "source of the sound" [29]. In analogy to the above considerations about Fourier–based approaches Gaver remarks that "the relevant attributes for everyday listening are not necessarily reducible to complex combinations of the attributes of musical listening".

Consequences and main points of focus

Given the well–founded outlines of the different fields of auditory perception, *everyday* and *musical listening*, the conclusions I draw for the work at hand are several. First, *everyday listening* can be considered as what I called a distinct channel of perception within the auditory mode. *Everyday listening* has its own attributes that form a domain different from that of attributes of music (at least in a traditional sense) or speech. The fact that identical acoustic stimuli can give rise to percepts of *everyday listening* or *musical listening* or contain a ver-

[11] Indeed it is exactly the central position of the user in its complex behavior not entirely covered by analytical theories that give auditory display a design aspect, instead of letting it appear as a pure engineering task.

[12] . . . i.e. an auditory material impression — I will come back to the subject later,

[13] . . . if we assume that the human brain is based on processes that do not generally contradict mathematical/physical formalization

[14] This is an example of where I see the term "sound attributes" as problematic; Gaver here refers to attributes derived formally from the acoustic signal.

bal message, compare e.g. a human voice, justifies the idea that these categories indeed are based on distinct perceptual mechanisms. As a consequence, a potential can be expected to convey information to a listener, in our context a user, by activating and exploiting capabilities of *everyday listening*, that might not or hardly be possible to transmit in other ways. The final evaluation (chapter 4) of one sound model (embedded into a larger tangible–audio-visual interface, the *Ballancer*, section 3.3) indeed validates this hope. I show that the information perceived from the (ecological, non–speech, "non–musical") sound is exploited in optimization of control movements, a phenomenon that is at least in that concrete appearance hard to imagine e.g. for verbal information (and in fact has never been proven before at all). It is actually this measured effect in gestural reaction through which the perception of ecological information is shown. The aspect of *continuous feedback and interaction* forms a main difference to the pioneering works of application of *everyday listening* by Gaver [29] [30] [31], who does not look explicitely at the immediate gestural exploitation of *continuous* sonic information (as it is usual in everyday surroundings). This difference is reflected by Gaver's term "*auditory icon*", standing for "caricatures of naturally occurring sounds", which implicates a rather closed, a–priori known unit rather than a reactive continuous dynamic behavior, as I aim at in the "*sound models*" [15] of chapter 2. In the test interface used for the evaluation, the *Ballancer*, users are seen to react continuously on the uninterrupted rolling sound which in turn continuously reflects the results of user input (the movement of the virtual ball).

When introducing his approach to use *everyday listening* for sound in computer interfaces [29], Gaver focuses on the question of mapping data (to be conveyed) onto sonic properties. He points out that in previous works of sonification (e.g. [11]) dimensions of data are represented by abstract attributes such as pitch or loudness, a strategy based on traditional understanding of sound and hearing; the latter still appears to be the norm (see e.g. [6]). Gaver's approach, as ours, instead is based on the mapping of dimensions of data to dimensions of sound sources, and he discusses in depth the strengths of this strategy. Without repeating detailed argumentations I only mention that the use of ecologically expressive sounds promises to reach better learnability through stronger *articulatory directness* [38]. The success in reaching this goal is another point in the evaluation chapter (4) where I show that subjects recognize the modeled scenario (rolling) from the synthesized sound alone and that the sonic feedback from the model allows to understand a larger control metaphor and its use (balancing a ball) without further explanation. It is seen that the sound model (of rolling) through its ecological expressiveness has a potential to steer a user's actions without dedicated explanation or training. This mapping of sounds addressing *everyday listening* and events in the real world, that people learn from early childhood in their interaction with the world, is in contrast to possible symbolic meanings of abstract sounds that have to be learned specifically. *Everyday listening* can be seen as our constant tendency and capability to

[15] The term and the connected aspects will be properly introduced and laid out below.

decode the "natural sonification" of events and interactions in our surroundings in environmental sounds.

Gaver also addresses the question of metaphors between sound-producing events and represented data and processes, that are necessary when a *nomic* mapping between a sound model (in his case: an *auditory icon*) and data is impractical. He states "the creation of overall metaphors from the computer interface to other environments (e.g. the desk-top model)" as a possible solution, and remarks that "such metaphors were created with predominantly visually-oriented systems in mind" and that "the addition of sound is likely to shape those metaphors in new ways". The *Ballancer* (chapter 3, section 3.3) forms one concrete (and thorough, evaluated) step towards such new audition–oriented or "audition–friendly" metaphors. It has to be kept in mind that, as explained at the beginning of this section, I do not start from a specification of data to be conveyed, since it is argued that new modes of interaction may enable the convection of data not yet imagined to be transported in a human–computer interface. I thus do not (yet) close the metaphor in a practical application (such as a steering task).

Finally, the mentioned restrictions of traditional theories and tools of psychoacoustics with respect to everyday sounds, raises the problematic of how to approach the modeling of such sounds. How can we develop sound generating algorithms that express ecological attributes if the established psychoacoustic methods and thus also conventional techniques of sound synthesis (that are based on these methods, e.g. subtractive, additive or FM synthesis) are not sufficient? The simple recording and playback of environmental sounds conflicts with the goal of continuous reactivity. A possible solution that is made use of here, is based on physical descriptions of sound sources and the main motivation and principle is shortly sketched. Auditory attributes of *everyday listening* are by their nature connected to physical attributes of sound sources, because that is the central function of ecological perception, to deliver information about physical objects and processes (or events) in our surroundings. If this connection is known sufficiently well and we have a satisfactory physical description of a sound emitting process, we may produce ecologically expressive sounds by predicting or numerically simulating a physical behavior. Gaver has partly applied this approach and noted the development of physical models as being "extremely difficult" and in fact some time has to be spent on according constructions and implementation (section 2.3). I refer to this general principle as "physics-based" and use the contrasting term "signal-based" for all methods of sound generation or psychoacoustics that approach questions of auditory perception starting from a description of the acoustic signal without considering the physical nature of its source. The latter term covers mainly the traditional Fourier–related theories and techniques. There is of course no reason to ignore such traditional psychoacoustic knowledge where it is applicable and helpful. E.g., *modal synthesis* will be used in the work of chapter 2, based on a particular, the "modal", description of vibrating objects, that is particularly well–suited when their acoustic appearance is of concern, because it relates closely to the traditional acoustic parameters. In chapter 2 I describe

more systematically an approach to develop sound models by exploiting and integrating both physics-based and signal-based techniques of sound synthesis. Of course, starting from a physical description is generally more demanding also in terms of computation, since the temporally evolving state of a physical object contains (usually much) more information (in a mathematical sense) than the resulting acoustic pressure signal at the eardrum. But this "overhead" may be the at times necessary tribute we have to pay to the human auditory system, that is admirably potent in gaining information from acoustic signals, which is all but obvious mathematically.

Relevant for the development of ecologically expressive sound algorithms are of course all those psychoacoustic works that examine ecological auditory perception, i.e. the attributes of *everyday listening*, "what people hear", and the possible formal reduction or connection of such attributes to properties derived from acoustic signals, "how we hear" them. Vice versa, physical models may serve as a tool to examine auditory perception (see e.g. [62]). [32] contains an annotated bibliography of *everyday listening*; I only briefly list the main works that were concretely important for the work described in chapter 2, details of application are given at the according places.

- Vanderveer's work [74] that has been mentioned, is important here in so far as it is the first application of the *ecological approach* to auditory perception and demonstrates the constant tendency and enormous potential of the human auditory system to extract information about sound sources from the signals arriving at our ears.

- Warren and Verbrugge's study of breaking and bouncing sounds [77] was the inspiration and basis for the modeling of "breaking" described in section 2.4.2. It is a striking demonstration of the perceptual relevance of "macro-temporal" patterns for the classification or recognition of acoustic stimuli.

- William Gaver's classic works [29] [30] [31] have already been mentioned and cited extensively, as they are immensely important for this thesis in many respects, which is reflected in the frequent use terminology introduced here (such as *everyday listening*). Gaver discusses in detail the psychological theories and viewpoints concerning the perception of environmental sounds [29], proposes a systematic for classification [29] and study [30] and develops the first attempts to exploit these notions in human–computer interaction.

- Wildes and Richards' examination on material perception [78] somewhat exemplifies an ideal result of psychoacoustics from the practical viewpoint: here, a widely valid and applicable connection is derived between physical properties of a sound source — (a coefficient of) "internal friction", mathematical parameters of emitted signals (a relation of decay times and the frequency of partials) and a perceptual phenomenon, namely the tendency/capability to estimate/recognize material classes (metal, glass,

wood...) from heard sound. These results have been recognized as highly useful and taken up and exploited repeatedly (see e.g. [40]) and are used explicitely as well as informally (in practical details that are not always documented) in the following.

- Freed's examination of "perceived mallet hardness" [27] points out the important phenomenon of auditory perception of hardness that gives weight (sections 2.3.1 and 2.3.4) to the respective potential of the algorithm of impact interaction used in chapter 2. Freed's results are not directly applicable as unambiguously e.g. as those by Wildes and Richards [78] (that concern material attributes) and the model of impact used later allows parallel achievements through a physics-based parameter without relying on signal-based parameters as those derived by Freed.

- A number of psychoacoustic works has addressed the question of auditory perception of shape (e.g. [42], [41], [45], [57]) but respective results do not (yet?) appear strong enough to form a reliable, manifest basis for the conveyance of information in human–computer interaction (which is the final scope in this work).

1.2 Scopes and terms

General subject of the work in chapters 2 and 4 is the development, implementation and evaluation of sound generation algorithms that can supply human–computer interaction with a sonic channel that exploits mechanisms of *everyday listening* in the spirit explained in section 1.1.

In comparison to Gaver's pioneering works [29] [31], I put explicit stress on the aspects of *continuous*, *dynamic* sonic feedback and in particular continuous reactivity, e.g. on a user's input, i.e. *interactivity*. These latter demands largely exclude the use of sample playback in technical realizations and are the reason why I do not stick to an adoption of Gaver's term "*auditory icon*" but use the one of "sound model" instead. Of course this introduced term may be criticized as hiding behind the wide range of possible meanings of "model" and, perhaps more important, because it is *not* simply sounds that are modeled but sound emitting scenarios or configurations of sound sources. "Sound cartoons" as a spontaneous alternative term on the other hand somewhat conflicts with the aspect of interactivity (cartoons are usually not interactive...) and "sound scenario model", which might be the most fitting name for the implemented algorithms, appears rather edgy and unhandy. It has to be noted that Gaver's *auditory icons* do absolutely not exclude continuous interactivity in so far as they are based on synthesis and not necessarily use stored samples. The term "icon" however, that implies a rather fixed, static character, reflects the fact that continuous reactivity/interactivity is here not (yet) one of the central scopes (but rather the exploration and use as such, of *everyday listening*). Of course these shifted weights reflect the time span of 15 years that has passed since Gaver's first works: the second main point of focus in this thesis is the possible

exploitation also of techniques and experiences of sound synthesis that have become available or practicable only in those recent years.

In contrast to other related works of sound generation that make use of physics-based techniques, I consequently take into account and take further the noted difference of *ecological auditory expression* and straight, possibly realistic, imitation. In order to achieve clearness of expression, it is generally preferable to stress and possibly exaggerate or isolate certain attributes of a sound source, on the cost of others that are considered of minor importance. I call this process of (auditory) *caricature* [16] of "real" scenarios "*cartoonification*".

Finally, in confrontation to works of sound synthesis as a goal per se, flexibility and ease of control and in particular economy of computation and implementation are other major issues of the following work: the audio channel of an interface within a larger system can naturally not demand the same amount of computational power and attention during implementation and tuning as a stand–alone sonic application, e.g. an electronic musical instrument. Economic implementation can often be achieved in parallel with, or as a consequence of, *cartoonified* — simplified and abstracted — expression; just like graphical icons or cartoons, that are both, easier to draw (cheaper in "implementation") and clearer to comprehend in their meaning, than photorealistic pictures or films.

The way the described goals are achieved is by integrating various techniques of sound synthesis from state–of–the–art physics-based to rather conventional signal–based and I try to extract and generalize the essence of the ideas and experiences in sound modeling in a "concept of sound design" (section 2.1). It is understood that I use the term "sound design" not as restricted to its most familiar meaning as associated with film production, but more widely in the sense of designing, "constructing" or shaping, sound or sonic appearance, to a user's benefit. Concretely tools are provided that allow the conveyance of information through mechanisms of *everyday listening*; a further use of the developed sound models, e.g. in the context of music or in a central position of a sound-based game, is of course not excluded (but welcome...).

The cycle that started from psychoacoustic impulses and leads into questions of sound generation and technical details, is closed by an evaluation (chapter 4) that demonstrates the success in reaching the initial goals and demands. The results of the evaluation in turn justify and approve the expectations explained in section 1.1 and contribute to psychoacoustic knowledge.

[16] The sound models introduced later are meant to be characteristic and clear, just as graphical caricatures; and as in the graphical case this does often *not* imply realism but rather simplification and exaggeration.

Chapter 2

Sound modeling

2.1 A *hybrid (sound design) architecture*

The central principles of *ecological expression*, *cartoonification* and (continuous) realtime *reactivity* for a design of auditory display have been explained and reasoned for in chapter 1. The static, repetitive character of *sample-based* sound has been mentioned, and its general incapability to reflect dynamically ongoing processes and actions, such as user input. On the other hand, it has already been motivated in chapter 1, that sound synthesis by *Physical modeling* [39][13], an approach that has reached increasing popularity and impressive progress in the last decades, relates naturally to the first scope of *ecological expression*. This approach starts from a physical description of a given system in its possible temporal behavior, i.e. generally a set of partial differential equations, rather than an analysis, or simply recordings, of typical emitted acoustical signals. In a final computational realization, or "simulation", the control parameters are exactly the chosen physical variables. The expression of ecological attributes should on this basis be straightforward (at least as long as the dependency of these attributes on physical values is sufficiently known [1]); it is not necessary to make a connection to, or even be aware of, properties of the acoustic signals that are finally generated [2]. In particular — probably one of the most striking arguments for the use of physics-based algorithms here —, in achieved acoustic results, ecological/physical attributes may be conveyed (to the listener), whose auditory perception is not (yet) adequately explained in signal-theoretic terms. For example, the physics-based model of impact interaction (2.3.1), that is one core element of the following sound models, can produce complex transients that reflect properties and the state of the involved objects and attributes as forces

[1] This remark is not completely marginal: e.g. would probably most people agree to have some intuitive idea of the "hardness" of familiar objects, but an exact physical description of this intuitively used general attribute is not as trivial. In particular may a physical definition of hardness perhaps not completely overlap with the everyday use of the word.

[2] ... apart from the fact that essential principles of digital audio and discrete–time implementation, such as the Nyquist theorem, must be understood and respected.

and velocity. Such transients have been found of high perceptual relevance (see e.g. [50]), but a satisfying theory about their perception is currently not available, and accordingly no signal-based [3] method for their detailed dynamical synthesis. *Physical modeling* also supports the central goal of *reactivity*, since the involved physical variables, i.e. the control parameters, can be changed dynamically (in an appropriate realtime implementation). The auditory result may vary dramatically with such changing control input. This is in strong contrast to sample-based sound, where variations of the few free parameters of filters and envelopes, usually can not overcome the static, repetitive character of used fixed sound samples.

From a theoretical, physicist's standpoint, an ideal model might be equivalent to the most advanced, most detailed physical description of the scenario in question. Mechanical (or electro-mechanical) systems are generally described in terms of (partial) differential equations, and the temporal evolution of such systems, including their acoustic appearance as one aspect, is, in a straightforward approach, found by numerical (i.e. usually discrete–time) solution [4] of the underlying equations. Applying this straight approach — discretizing the most thoroughly describing equations — to complex mechanical scenarios generally results in complex algorithms, that are computationally demanding and highly specialized. For example, the implementation of the complete equations of a falling and bouncing object, in its macroscopic and inner behavior, in three dimensions at a standard audio-rate (e.g. 44.1kHz), would be unaffordable to be used interactively in realtime in our context of human–computer interfaces. Also, it would not be possible to expand the same algorithm for a sound-model of a physically rather different scenario such as "breaking"; the whole process of development, starting from different equations, would have to be repeated. [5]

The standpoint in this work is different from the "theoretical, physicist's" one, in that we are not interested in a system per se as a whole, but only in one partial aspect, its acoustic appearance, or, more precisely, its perceptual, auditory impression. Yet more important to keep in mind, human auditory perception is at the center of interest here, not "simply" in the sense of trying to match as close as possible [6] the auditory pictures of model and "real thing", but through the auditory conveyance of *ecological attributes*. An auditory impression however is something else than the sum of contained *ecological information*. To understand this last reflection, one may recall the common experience that the voice of a friend sounds different on the phone than in direct face–to–face conversation; but it is not necessarily clear in how far (if at all) both auditory impressions differ in terms of the *ecological information* that they transmit? Finally, *ecological attributes* are of interest as a mean to represent various information [7] in human–computer interaction, thus giving individual

[3] In chapter 1 I have explained how I use the term "signal-based".

[4] ... if we exclude the rare cases of analytically solvable systems,

[5] This example is given here for comparison with the derivation of sound models of "bouncing" and "breaking" described in the following sections (2.3 and 2.4)

[6] ... following whatever norm of comparison,

[7] It is not the scope of this thesis to thoroughly organize the possible types of information

"weights" to such attributes; this is the essence of *cartoonification* (1.2). As a consequence, I also make use of a "*model-based*" approach to sound synthesis, however putting raised stress on three of the possible aspects of the term "model": simplification, representation and abstraction. "Simplification" here ideally refers to the complexity of both, computation and structure/handling. "Abstraction" aims at the auditory appearance of the model, in confrontation with a "real" mechanical pendant [8], that may be more generic, less concrete, less "natural", more "artificial", as well as its internal structure. As an example for the latter, section 2.4 will present a sound model of "bouncing" that also covers "breaking", thanks to abstractness in its inner structure. The following sound models acoustically "represent" typical scenarios without necessarily reproducing or imitating them (or the emitted signal).

The practical implications of this conceptual "position-fixing" and the precedent general considerations about *physical modeling* are as follows. Physics-based algorithms in the straight sense, i.e. based on the numerical solution of describing differential equations, are used where it is meaningful, i.e. offers clear advantages of the nature given above, and affordable (concerning implementation). Section 2.3 describes two such directly physics-based models, a general vibrating object in *modal description* [9] and an algorithm of impact interaction. Instead of expanding this straight approach to larger, more complex scenarios in a "brute–force" strategy, I use more abstract structures to cover higher–level processes as described in section 2.4. At this stage signal-based approaches are integrated that remind of older techniques of sound synthesis; e.g., signals of idealized waveforms are used, sine, saw or noise signals. Typically, these higher–level structures make use of, and control, a straight physics-based, lower–level audio core. Often, this reflects an analog structure in the sound and its causing event; "bouncing", "dropping" or "breaking" scenarios e.g., contain single impact events. Abstraction however, already starts at the lowest level where simplicity for describing equations is preferred over detailedness. The impact model below (section 2.3.1), e.g. is one-dimensional and all micro-contacts in modeled scenarios are reduced to this one dimension. Such perception-oriented hierarchical structures, integrating modeling processes at different levels, from straight physics-based to more abstract, and also exploiting signal-based techniques, are summed up under the term "hybrid architecture". A practical instantiation of the general concept will be described in the next section (2.2), at the field of sounds of contacting solid objects. The suitability for, and success in, reaching the initial scopes of *reactivity* and *cartoonified*, informative *ecological expression* is finally proved in the chapter of evaluation (4).

---

that may be desirable to communicate to a user through non–speech sound. What is done here, is to try and "learn from nature" and build (hopefully powerful) "communication roads".

[8] Of course the sound of the models at hand is not "abstract" in an absolute sense, e.g. as a sine tone, but relative to direct "real" sounds.

[9] The theory behind this term is shortly summarized in its properties and relevance in this thesis in the respective subsection (2.3.2).

2.2 Modeling of contact sounds

The perception of ecological attributes like material or shape from sounds of contacting solid objects is common experience. Probably everybody can comprehend from everyday experience that a struck wine glass sounds "like glass" and will (from its sound, without seeing it) not be taken for a wooden object, just like a dropping spoon will be auditorily perceived as being of metal and a bouncing ping pong ball will not be confused with a dropped plastic bottle. Such or similar examples of *ecological perception* have been examined in a relatively large number (compared to other classes, such as fluid or gas sounds...) of psychoacoustic studies (chapter 1), that realize and signify the importance of contact sounds for *everyday listening*. In fact, from the viewpoint of auditory perception of ecological information in our everyday surroundings, scenarios of contacting solid objects form probably the most important class of all familiar sound emitting processes. Accordingly, also for sound synthesis, contact scenarios have been recognized as a crucial subject in several works [71][73][62]. However, these previous studies either focus on the resonance behavior of involved objects and widely neglect the important transient states of the interaction, or follow an expensive, straightforward "brute–force" approach. In the first cases [71], [73], only the description of the resonating objects is physics-based; these are described in the modal formalism [1] that also plays a role in this work. For the interaction itself, fixed force-profiles are assumed, which can be seen as a sample-based technique on a different level. Indeed, the whole resulting model is of a source–filter structure, that ignores the dynamical nature of individual impact transients, and gives less convincing results especially in cases of frequent or continuous contact such as "sliding" or "rolling". Other works (see e.g. [9][8]) are based on the numerical solution of possibly detailed equations describing the complete three-dimensional objects and their interaction, which may lead to highly realistic simulations. Implementations become accordingly expensive (in terms of computation and control) and do not fulfill the pretensions that are at the center of this thesis. From this situation and the reasons just sketched, sounds of contacting solid objects appeared as a particularly worthwhile subject for the practical application and approval of the general sound design concept. Besides, contact scenarios lend themselves well to apply and demonstrate the hierarchical, hybrid modeling architecture outlined in section 2.1: many examples of contact can be deconstructed into events of "micro-contact" and acoustically less significant phases. E.g., from the scenario of a dropping object single impact events can be isolated and the global (rebouncing, falling and turning) movement can be accounted for separately. Other contact scenarios, such as "squeaking doors" or "rubbed glass", can be based on friction interaction; a closely related work deals with this complex [59]. [10] In the following, the development of several sound models based on impact interaction is described. Besides "bouncing" and "dropping", it is seen that also "sliding" and "rolling" and even "breaking" can be conceived and

[10] An implementation of a friction model described in [61] uses some of the technical structures developed as part of the work presented here.

modeled in this way. At the lowest level in the modeling hierarchy, a physics-based (in the closer sense) algorithm of impact interaction is developed and implemented. More abstract, perception-oriented, higher–level structures that take account of "macroscopic" geometrical aspects, make use of, and control, this central audio core. Figure 2.1 gives an overview of the implementations and some integrated interaction examples as presented in chapter 3.

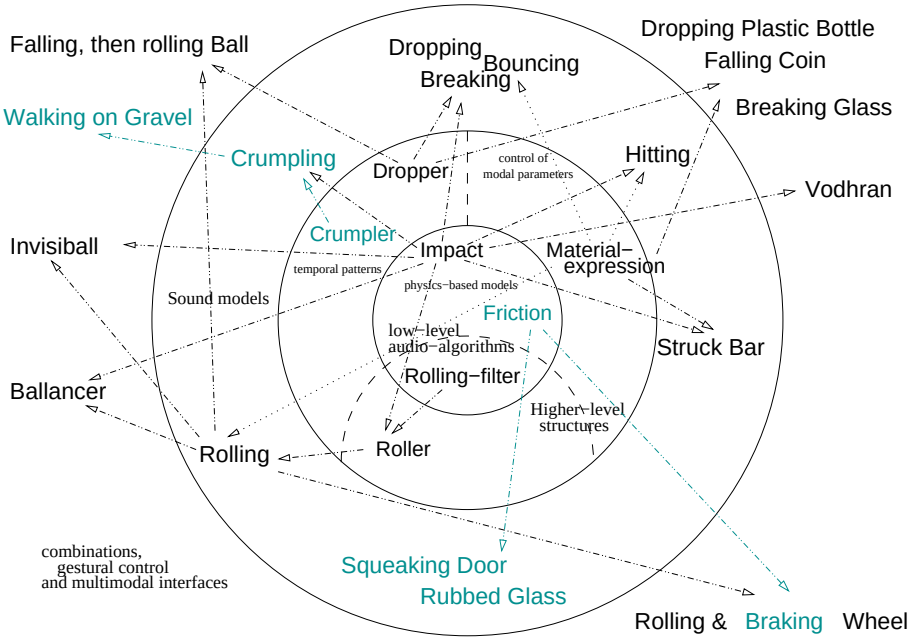


Figure 2.1: Overview of real-time sound models of contact scenarios and their underlying structures, as developed during the course of the SOb project. The graphical layout in nested (circular) fields reflects the structural hierarchy: physics- and geometry-based (dashed half-circle) low–level audio algorithms in the center, completed with surrounding higher-level objects, resulting sound models in the largest circle and finally combining example-scenarios and multimodal interaction examples. Arrows indicate dependencies and are to be read as "contributes"/"used to realize". Among the audio core algorithms (inner circle) the *Rolling-Filter* differs from the straight physics-based models of impact and friction, in that it reduces macroscopic geometric features to the parameters of microscopic interaction (section 2.4.3); this special aspect of the rolling model is indicated by the dashed half-circle. The models represented in grey are not results of the work of this thesis but make use of results developed in the following; such dependencies are indicated by the arrows or, in the case of the friction *module*, explained later (section 2.3.3).

2.3 Low–level physics-based modeling

2.3.1 A physics-based model of impact interaction

At the heart of the following work on contact sounds stands a model of impact interaction, according to the general consideration that the scenarios looked at in the following, "hitting", "bouncing", "rolling"..., can be based on microscopic impacts. This impact model is physics-based in the close sense described in section 2.1, i.e. based on a mathematical equation describing a physical process. As an advantage of this choice, with respect to previous works, complex dynamical sound results are obtained that reflect and express a wide range of attributes of the modeled scenario, also beyond the current range of signal-based theory and methods. Other works that have realized the central position of impacts for ecological perception [73][71] focus on the resonance, i.e. decay, behavior of the involved objects and widely ignore, or only roughly depict, the transient stage of the event. For the interaction phase of the objects, fixed (impulses, semi-cycles of cosines) or statistic (noise-impulses) force profiles have been used, that are only slightly adaptable to physical/ecological attributes, and do not individually react on the current state of the objects. [11] This practice reflects the current state of knowledge about the perception of impact transients: while their high perceptual relevance is recognized, only few signal-theoretic indexes have been derived that roughly describe qualities of ecological perception [27], but an exhaustive closed theory is (currently) not available. Using a physics-based description not only for the resonating contacting objects (as has been done in depth in preceding works [71][73][19]) but also for the interaction itself, we are not restricted by the limits of theories of perception and of acoustic signals.

However, already the physics-based impact model includes a degree of abstraction that implies efficient implementation as well as adaptation to a wide range of concrete situations. A one-dimensional term of interaction force f is used, that depends on an, as well one-dimensional, "distance variable" x. The three-dimensional local geometry of both interacting objects is only represented through one parameter, α, and possible simultaneous interaction in other directions is not taken into account at this stage. This leads to a compact efficient algorithm that strikes the main interaction properties.

$$f(x(t), \dot{x}(t)) = \begin{cases} k(-x(t))^{\alpha} + \lambda(-x(t))^{\alpha} \cdot (-\dot{x}(t)), & x < 0 \\ 0, & x \geq 0\ . \end{cases} \tag{2.1}$$

Here, 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. For a positive distance, $x \geq 0$, the two interacting objects are not in contact and consequently no interaction force occurs, $f = 0$. Negative distance values, $x < 0$, mark the case of contact, i.e.

[11] These techniques would thus not be usable for the higher–level models below, e.g. of rolling, as will become clear.

resulting deformation and a corresponding (non–zero) interaction force f. Similar formulas describing contacts of solid objects and the effective contact force have been used for sound synthesis before, mainly for the simulation of piano tones (e.g. [33], [69]). The equation (2.1) used here originates from a work of robotics [49], i.e. was not originally derived with the aim of generating sonic feedback. Its adaptation for sound generation was suggested and numerically solved by Avanzini, Rocchesso et al. [4][2]; [60] contains a detailed discussion of the origin and context of the equation (2.1) and discusses its relationship to other similar describing formulas. Related previous works in sound synthesis combine interaction terms like the one used here with resonators described as *digital waveguides*. This common technique is advantageous in many aspects and widely used for the synthesis of musical instruments such as string or wind instruments with their (basically) harmonic spectra, but less suitable for the modeling of resonators with inharmonic spectra like most everyday objects. A thorough combination of an efficient, dynamic physics-based model of contact interaction of the type of equation (2.1) with fully general modal resonators (as explained in the next section 2.3.2), in theory and implementation, for the aim of modeling everyday sound scenarios has not been established before and forms a new contribution to the field of sound synthesis. The computational implementation of the model, integrating interaction and resonators in a modular fashion is described in section 2.3.3.

Giving more weight to computational economy under the premise of *cartoonification*, an alternative simplified equation for the interaction force f is suggested and has been implemented:

$$f(x(t), \dot{x}(t)) = \begin{cases} -(kx(t) + r\dot{x}(t)), & x < 0 \\ 0, & x \geq 0\,. \end{cases} \tag{2.2}$$

This linearized version is received from (2.1) by setting $\alpha = 1$ and ignoring the factor of $x(t)$ in the second summand, i.e. replacing it with a constant. The derivation is directly based on experiences with the acoustic results gained from (2.1) and considerations of implementation (i.e. computation and control): the term of (2.2) for the contact force is (piece-wise) linear and thus particularly easy to solve numerically [12]; on the other hand, the experience with (2.1) showed that the dispersive term accounting for friction loss can be important in sound modeling while it is in its form in (2.1) delicate to handle. In comparison to other previously used formulas (see the overview given in [5]) (2.2) contains a linear term of energy dispersion ($-r\dot{x}(t)$) and originally accounts for the aspects of acoustic results and practical realization. Also known from the spring force of a damped harmonic oscillator, (2.2) is less detailed and rich in its acoustic potential than (2.1), but slightly more economic; it is thus a useful alternative for situations of implementation where cost of computation is particularly important and acoustic detail might be preferable to trade for practical affordability.

[12] Section 2.3.3 deals with this aspect.

2.3.2 Modal synthesis

Generally, several techniques could be thought of for describing the resonating contacting objects in the scenarios modeled here. Differential equations in three spatial dimensions may be solved "with brute force" numerically, i.e. by spatial discretization over the whole domain of the object, which generally leads to algorithms that are highly expensive for computation in realtime in an interface context. *Digital waveguides* as an efficient alternative for resonators with basically harmonic spectrum (wind or string instruments...) are not easily adapted to more general resonators, such as many everyday objects, without loosing its characteristic advantages (of efficiency in implementation)[19]. In fact, the impact model is implemented in a way that allows the easy integration of very different resonators (as explained in section 2.3.3). For the remainder of this work, I use an approach that is, besides being economic in implementation, advantageous in our context of steered auditory perception in many aspects.

"Modal Synthesis" is based on the description of a resonating object in coordinates that are not displacements and velocities (or other physical state variables such as flow/pressure) in the spatial domain of the object, but in terms of its *normal modes*. The state of the object is here written as the vector of its modal states, i.e. displacements and velocities along the axes of its modal coordinates, as it is in the straightforward spatial description seen as the vector of states of its spatial points or components. Figure 2.2 [13] sketches an idealized circular membrane, deformed from its planar rest position in two isolated modes. Modal and spatial description are in principle equivalent and related via a linear

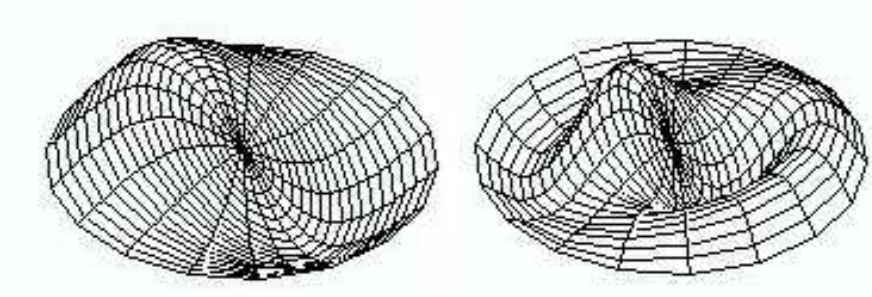


Figure 2.2: A circular membrane displaced from its rest position along the axes of mode(1,1) (left) and mode(1,2) (right).

basis transformation. The 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 by transformations of a matrix connected to the finite system of differential equations describing this finite case. While lacking the unmediated concrete (from the

[13] Thanks to Dan Russell of Kettering University Flint for his permission to include these pictures.

visual viewpoint) meaning of spatial state variables, the modal formalism is a powerful (standard) tool for the examination of vibrational movement [14], and particularly suitable under our premises. The main points are given in the following, exact derivations and details can be found in dedicated literature (e.g. [28]).

Modes are of countable number and therefore discrete, in the distribution of their frequencies and shapes. [15] This property is of high importance for implementation and simplification, because, due to the finite bandwidth of the human ear and practical "thresholds of significance" in frequency and amplitude, the number of modes to be considered in practice is always finite. The inner (i.e. without external feedback) temporal development of the state of a given system along each modal axis is independent of its state and development along the other modal coordinates. The differential equation describing the system splits into a series of independent equations, the modes are *decoupled*. If we denote by x_j the displacement of an object along the axis of one mode indexed by j, the modal state $\mathbf{w}_j = \binom{x_j}{\dot{x}_j}$ follows an equation of the form

$$\ddot{x}_j + r_j \dot{x}_j + k_j x_j = f_j, \tag{2.3}$$

where $r_j \geq 0$ and $k_j > 0$ are the damping and the elastic constant for this jth mode, respectively, while f_j is the sum of external forces acting on the mode. Equation (2.3) is known as the one of a damped harmonic oscillator, and solvable analytically; we see that a (one-dimensional) harmonic oscillator is exactly a system of one normal mode. The corresponding impulse and frequency response of one mode, i.e. of a system excited by a force along one modal axis [16], are known from the theory of the harmonic oscillator. For sufficiently small damping, $r_j^2 < 4k_j$, the impulse response $h_j(t)$ of (2.3) is given by

$$x_j(t) = h_j(t) = e^{-t/t_j} \sin(\omega_j t)\ . \tag{2.4}$$

We see that the free (i.e. without or after any external influence) resonance movement of one individual mode is rather simple — from the standpoint of its auditory perception—, an exponentially decaying sinusoid of a fixed frequency [17]. The *modal frequency* ω_j and the decay time t_j are given by

$$k_j = \omega_j^2 + 1/t_j^2, \quad r_j = 2/t_j\ . \tag{2.5}$$

Again, for sufficiently small damping the resonance frequency is approximated by $\omega_j \simeq \omega_j^{(0)} \triangleq \sqrt{k_j}$.

The resonance behavior, i.e. the frequency response, corresponding to (2.3) (the Fourier transform of (2.5)) is that of a lowpass filter with a peak around

[14] Modal analysis is also one of the standard techniques used in control of the vibration of mechanical systems in industrial, e.g. car, design.

[15] It will become clear in the following what are the "frequency and shape of a mode".

[16] The different excitation of various modes, e.g. through mechanical interaction at different spatial points, is discussed below.

[17] ω_j depends only on the mode, thus "*frequency of the mode*".

this mode (or resonance) frequency. The bandwidth of this peak is proportional to the inverse of the mode's decay time.

The basis transformation between the modal and spatial state variables is linear. Concretely, the position–velocity configuration $\mathbf{w}_P = \binom{x_P}{\dot{x}_P}$ in a specific "pickup point" [18] P is a weighted sum of the mode states $\mathbf{w}_j$; conversely, an external force f input to the system at P is distributed to the distinct modes with the same (position dependent, indicated by the subscript "$_P$") weighting factors.

$$\mathbf{w}_P = \sum_{j=1}^{n} a_{P\ j}\ \mathbf{w}_j\ ,$$

or equivalently:

$$x_P = \sum_{j=1}^{n} a_{P\ j} x_j = \mathbf{a}_P \mathbf{x} \quad \text{and} \quad \dot{x}_P = \mathbf{a}_P \dot{\mathbf{x}}\ , \tag{2.6}$$

where $\mathbf{x} = (x_1, \ldots, x_n)'$ is the vector of the modal position variables as in (2.3), and $\mathbf{a}_P = (a_{P\ 1}, \ldots, a_{P\ n})$ are the weighting factors at P. Vice versa

$$f_j = a_{P\ j}\ f, \qquad j = 1, \ldots n \tag{2.7}$$

with f_j as in (2.3). The transfer function connected to a pair P, Q of points on the object, i.e. the resulting movement picked up at Q caused by a force applied at P, is a weighted sum of the transfer functions of the single modes, with weighting factors $a_{P\ 1}a_{Q\ 1}, \ldots, a_{P\ n}a_{Q\ n}$. In other words, the impulse response of the whole system to an impulse at P as measured at Q is the weighted sum of exponentially decaying sinusoids as in (2.5). The according frequency response is the weighted sum of resonant lowpass filters, a "filterbank".

From the independence of the modes and the linearity of the transformation between modal and spatial description, the temporal movement corresponding to the example states in figure 2.2 is seen to be characterized in the following way: all points on the membrane perform a sinusoidal movement around their middle position, "swinging" up and down periodically, with the fixed frequency of the mode and exponentially decaying amplitude. Due to the linearity of the coordinate change, at any instant the membrane forms the same shape (therefore: "shape of the mode"), just "scaled" or "stretched" perpendicular to the rest plane. All points move *in phase*, i.e. pass the central rest position simultaneously (as long as only one modes is excited). In particular, points on the section lines of the mode shape and the rest plane (see figure 2.2) do not move at all; on these *nodes*, the mode can not be excited nor "picked up" (measured). The general free movement of the membrane is a superposition of such single–mode movements; in the general case, the spatial points of an object do not move in phase, as a consequence of the variable (depending on the point) weighting of modes.

[18] A mechanical pendant is an electromagnetic pickup, e.g. in an electric guitar, giving the sound of the movement of a string in "one point", i.e. a very small range.

The properties briefly lined out above and the main consequences from the viewpoint of this thesis shall be summarized in plain words:

- The modal approach is very general. A wide range of very different systems, from threedimensional solids, two- and onedimensional structures (such a membranes or strings) to gas-filled cavities, can be satisfyingly characterized in this way. *Modal synthesis* thus supports very well the goal of flexibility and generality (as opposed to specialization).

- The complete system, in its temporal behavior, is represented by a compact set of parameters: modal frequency and decay time for each mode, and a series of weighting factors (for all modes) at each possible point of interaction, i.e. of interest for (force) input or (e.g. sound) output.

- The equation of each mode can be efficiently implemented. Each mode takes the form of a second–order filter and various numerical implementation strategies exist, focusing on different aspects, such as efficiency, stability.... [19]

- Most important, the parameters of the modal description relate more closely to human auditory perception than a straightforward spatial description of a system. Sinusoids are among the most studied stimuli in psychoacoustics [50][53][54]. The corresponding view of the resonating object as a filterbank (in the case of given force input signals) of parallel resonant lowpass filters, also has an intuitive significance, e.g. roughly comparable to a parametrical equalizer. This finally shows the immediate (acoustic) perceptual significance of the parameters of the modal description that is gained in trade for the missing ostensive meaning of the modal coordinates themselves. [20]

- A very important consequence in our context is the potential to introduce in the modal description, extensive but well-directed and -controlled simplifications. 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 auditory 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*. The effects of reduction and simplification of the combination of modes have been studied extensively [72]. More general rules for the tuning of modes, e.g. in connection with the intended expression of material properties [78][40] exist that can be used for rather generic modeling than concrete simulations. In the direct spatial representation in contrast, the

[19] In the implementation here a state-variable formulation is used for the convenience of immediate access to position and velocity values; this choice might be reconsidered and compared, and possibly slightly optimized in terms of computational economy.

[20] It might be stated that the spatial description of an object rather refers to its visual appearance whereas modal properties have a closer relationship to auditory perception.

auditory effects e.g. of a reduction of the resolution of a finite-element discretization are usually not as clear and predictable a priori.

- Many techniques exist for the extraction and tuning of the modal parameters, starting from measured or recorded mechanical responses/sounds or mathematical analysis. Strategies and results range from very detailed and exact to rather rough and approximate. Due to the clear auditory interpretation, modal parameters may even be set completely intuitively for rather abstract representations.

2.3.3 Integration and implementation

The potential of the modal description as described in the previous section (2.3.2) for the modeling of ecological sounds has been recognized and exploited practically, so far mainly in a feedforward, source–filter [71][19] manner. [21] Models for the occurring force in impact interaction, similar to the one presented in section 2.3.1 have been used for the synthesis of musical instruments, namely the piano (e.g. [69], [33]). In the latter case, involved resonating objects are usually represented in the form of *digital waveguides*. A combination of the modal description with a fully three-dimensional simulation of the interaction, via a detailed reconnection to the spatial appearance of the contacting objects, is probably amongst the most powerful approaches developed so far [9]; yet, here the resulting complexity of computation and control is still remarkable and not suitable for the scopes of this work. The integration of the modal formalism with a physics-based, yet abstracted, efficient formula describing the interaction, is a new contribution of research work. Rocchesso et al. [4] describe the connection and numerical implementation of the interaction equation (2.1) with one of the contacting objects (the "hammer") being a free point-mass, hitting a damped harmonic oscillator as the second resonating object. The expansion of this model to the case of two interacting objects in fully general modal description [22], and the practical implementation in a modular structure, a general framework for the integration of different types of interaction as well as involved objects, is presented in the following.

Integration of impact interaction and modal resonators

In the modal description, the state of each of the two contacting resonating objects is seen as the vector of the states of modes, $\mathbf{w} = (x_1, \dot{x}_1, \dots x_n, \dot{x}_n)'$ (in the notation of equation (2.3)). Equivalently, with a simple reordering of rows,

[21] The strategy of using modal resonators with preassumed fixed force-profiles has already been discussed in the introducing paragraphs of previous sections (2.2, 2.3.1); other aspects of the same previous works, concerning the handling of modal parameters, are described and picked up in section 2.3.5.

[22] As already noted, a damped harmonic oscillator is the special case of a system with exactly one mode.

we can write

$$\mathbf{w}^{(1)} = \begin{pmatrix} \mathbf{x}^{(1)} \\ \dot{\mathbf{x}}^{(1)} \end{pmatrix} \quad \text{and} \quad \mathbf{w}^{(2)} = \begin{pmatrix} \mathbf{x}^{(2)} \\ \dot{\mathbf{x}}^{(2)} \end{pmatrix}, \tag{2.8}$$

where $\mathbf{w}^{(1)}$ and $\mathbf{w}^{(2)}$ are the states of the two objects and $\mathbf{x}^{(1)}$ and $\mathbf{x}^{(2)}$ the according position vectors in modal coordinates, as in equation (2.6). The interaction of the two objects is now described through a force term f of the type of section 2.3.1 (equations (2.1) and (2.2)), that only depends on some values $\mathbf{w}_P^{(1)} = \binom{x_P^{(1)}}{\dot{x}_P^{(1)}}$ resp. $\mathbf{w}_Q^{(2)} = \binom{x_Q^{(2)}}{\dot{x}_Q^{(2)}}$ that represent the relevant configurations of the objects in the form of equation (2.6), and acts on the two objects as described by equations (2.3) and (2.7). Naturally, $x_P^{(1)}$ and $x_Q^{(2)}$ here thought of as displacements in one direction at one *point of contact* P resp. Q, and of f as acting on the two objects in the same single point and depending on a distance value $x = x_Q^{(2)} - x_P^{(1)}$ via $f = f(x, \dot{x})$ as discussed in section 2.3.1. The variables and equations may however represent a more general situation; at this point I already use a formulation that enables the wider generality, extensibility and modularity of the final implementation that will be presented below. If we assume for convenience at the moment, the absence of other external forces than that of interaction, the temporal behavior of the entire system of both interacting resonating objects is determined by the following set of equations, that follow directly from equation (2.6) and the last remarks:

$$\ddot{x}_j^{(1)} + r_j^{(1)}\dot{x}_j^{(1)} + k_j^{(1)}x_j^{(1)} = \quad a_{P\,j}^{(1)} f(x, \dot{x}), \qquad j = 1, \ldots, n^{(1)} \tag{2.9}$$

$$\ddot{x}_j^{(2)} + r_j^{(2)}\dot{x}_j^{(2)} + k_j^{(2)}x_j^{(2)} = \quad a_{Q\,j}^{(2)}(-f(x, \dot{x})), \quad j = 1, \ldots, n^{(2)} \tag{2.10}$$

$$x = \sum_{j=1}^{n^{(1)}} a_{P\,j}^{(1)} x_j^{(1)} - \sum_{j=1}^{n^{(2)}} a_{Q\,j}^{(2)} x_j^{(2)} = \mathbf{a}_P^{(1)}\mathbf{x}^{(1)} - \mathbf{a}_Q^{(2)}\mathbf{x}^{(2)} \tag{2.11}$$

Here, $f(x, \dot{x})$ is one of the terms (2.1) or (2.2), but the following procedures are valid and applicable also under more general preconditions, i.e. for f being of a different form; this remark will be concretized later. The negative sign for f in the equations (2.10) reflects the fact the force acts on the second object in opposite direction (than on object 1). (This "−" sign, together with the one in the following term (2.11), might be omitted for convenience simply by inverting the weighting factors for the second object, $\mathbf{a}_Q^{(2)} \leftrightarrow -\mathbf{a}_Q^{(2)}$; but this simplification of notation would blur the logical structure of the implementational realization.) $n^{(1)}$ and $n^{(2)}$ are the numbers of (considered) modes of object 1 resp. 2.

Modular implementation

For a practical implementation, equations (2.9) – (2.11) have to be discretized in some way, i.e. transferred into a discrete–time recursive numerical algo-

rithm that can be executed in realtime by a computer. [23] During the process of discretization, particular attention has to be payed to avoid the occurrence of instantaneous feedback loops in the resulting algorithm. Instantaneous cross-dependencies of discrete–time variables make the recursive algorithm non-computable. [24] The avoidance of non–computable instantaneous loops is particularly critical if F is a non-linear function as in our case. For such non-linearities, Borin et al. [12] have developed a technique (the "*K-method*") to convert continuous–time equations such as (2.9) – (2.11) into discrete–time algorithms under prevention of non-computabilities. During the process of discretization and its preparation, the principle structuring of the system into resonating objects and a description of the interaction, in the above case two modal "*resonators*" [25] and the interaction force f, is generally lost or blurred. As a consequence, the whole process of discretization and implementation has to be repeated if one of the involved factors, *resonators* or *interactors*, are exchanged. The exact conditions, formulations and derivation of the *K-method* can be found in [12]. In the implementation described in the following stress is put on keeping modularity, from the initial formulation of the system, through to the discrete–time algorithm. Exactly, parts of the system, here the objects in contact and the description of the interaction, can be developed and implemented independently and interconnected dynamically, under certain assumptions, through a particular mechanism of interconnection. The original formulation of the *K-method* does not deal with this point, and I use a somewhat parallel (or specialization of the) approach, but located on the discrete–time level. Figure 2.3.3 sketches the problematic that is explained and solved in the following. I here try to keep the formulation possibly simple and do not make any effort to specify the precise conditions on continuous–time systems, nor to reach maximal generality. It shall suffice here to cover those scenarios that we are directly interested in and reach modularity under these nearer circumstances; the approach can however surely be generalized and worked-out beyond the immediate application in this thesis.

Behind the goal of modularity in implementation lies the central consideration, that many sound emitting scenarios (in particular all those that I look at in this work [26]) can be decomposed into distinct objects with individual, independent inner behavior, that interact in a specific way. Relevant for the interaction, is usually only a limited configuration, not the complete internal state, of the involved objects, and, vice versa, the internal behavior of the objects can be characterized independently from external interaction. As an example (and this is our concrete field of focus), solid objects can interact in various types of contact, such as impact or friction, at different points. Internal

[23] In later applications, additional unpredictable "realtime" parameters are included into the equations (such as varying surface profiles influenced by user actions), so that analytical solutions (to these equations), even if generally available, would be of no use.

[24] This remark will get more clear and concrete in the following.

[25] Below I define clearly my specific use of the terms "*resonator*" and "*interactor*" in the remainder of this chapter.

[26] . . . with a certain exception of "breaking", that will however in a *cartoonification* process also be reduced to "well-behaved" atomic components,

properties of such contacting objects can be described in different ways (such as the modal description presented before), independently from the interaction, that does not induce permanent changes to the objects — at least in these cases of interest here. On the other hand, information is exchanged only via the objects' configurations in the areas of contact; the entire state of the objects can generally not be deduced from their behavior in one, or some, limited contact areas. A structure of implementation is therefore of interest that allows to develop independently, computational algorithms representing distinct objects and processes of interaction, and to freely connect such algorithms without the need of further adaptation. In the following, I refer to representations — of whatever nature, discrete–time (mostly) or continuous–time, of independent objects in the explained sense as "*resonators*", and representations of processes of interactions as "*interactors*". The term *resonator* here aims solely at the presence of some sort of memory, i.e. some internal state that is relevant for the subsequent, future behavior [27]); no general a priori specifications, e.g. concerning linearity, are given at this point. For simplicity, *interactors* are here

[27]This notion of an internal state is reflected through a differential operator in continuous time representations, while we have some temporally changing state vector ($\mathbf{w}$) in the discrete–time algorithms, with a "state–update" algorithm ($\mathbf{w}(n) \rightarrow \mathbf{w}(n+1)$).

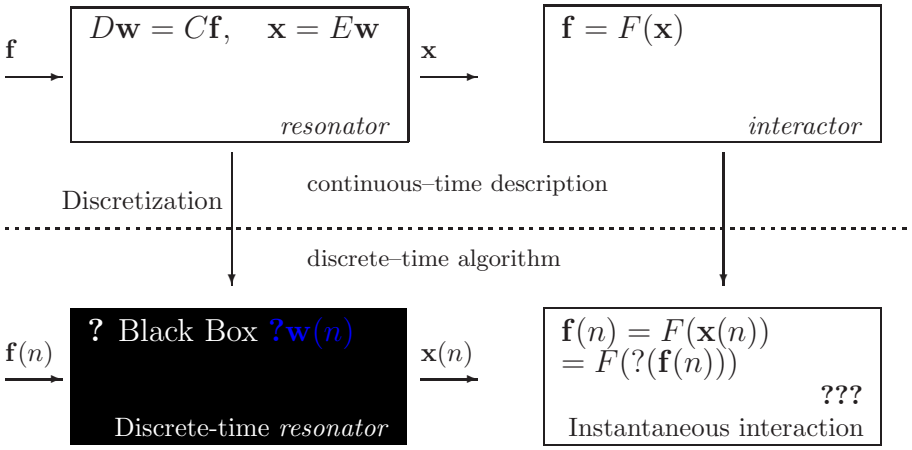


Figure 2.3: A sketch of the goal of "modularity" in implementation and the connected problematic. It is desirable to represent objects involved in modeled scenarios as "black boxes" that generate output from input values, without the necessity of further information (thus "black") about their origins or inner structures. This goal conflicts with the instantaneous cross-dependency of values, due to the description of the interaction. Such non-computabilities are usually excluded already at the continuous–time level (as done in the original *K-method*), which generally destroys the independence of *resonators* and *interactors*, that is the second main goal behind the term "modularity".

assumed to be memory-less, i.e. instantaneous relations; this assumption may be bypassed [28] but is unproblematic in the cases here and simplifies the description. *Resonators* and *interactors* are connected through input and output vectors that can most easily be thought of as forces $\mathbf{f}$ (coming from the interactor) and spatial position–velocity configurations $\mathbf{x}$, as in our concrete case; the complete state $\mathbf{w}$ of the *resonator* is generally not passed to the *interactor*. Figure 2.3.3 gives a sketch of the intended modular structure as described, and the connected problematic; only one *resonator* is depicted here, which has no influence on the validity of the following argumentations. Exactly, "modularity" here means, that discrete–time realizations of *resonators* and *interactor* formulas can be exchanged and "plugged" at this discrete–time level without any further knowledge about the internal algorithms or their origins, such as an underlying continuous–time model or the used technique of discretization (such as bilinear transform, Euler backward differencing...) whatsoever. Discrete *resonators* should be handable as "black boxes" that produce output vectors at every time step depending on their contemporaneous input vector and the hidden state-vector. It is seen that this goal conflicts with the instantaneous cross relationship given by the *interactor*: in figure 2.3.3, $\mathbf{f}(n)$ would be computed from $\mathbf{x}(n)$, which in turn depends on $\mathbf{f}(n)$; this loop can not be resolved without additional information about the internal structure of the *resonator*, i.e. without "breaking the black box". A non-computable instantaneous feedback-loop occurs.

The described problem is solved and modularity is reached in the development and interconnection of the *resonators* and *interactors* through a "labeled–black–box" approach. It is clear that discrete–time *resonators*, in the situation of figure 2.3.3, can not be handled as strict black boxes, in the sense of passing input to output vectors without additional information. However, as we will see now, it is on the other hand not necessary to reveal the origin and complete internal structure of the *resonator* algorithm, nor to reconstruct the whole computational structure for each change of objects or interaction. Under certain presumptions on the *resonator* algorithm we are able to resolve the instantaneous feedback loop, with the help of a "label" attached to the black box, representing minimum information about its hidden internal structure. The important point here is the exact specification of these presumptions on the *resonator* and of the "minimum information necessary", and the derivation of a uniform representation and interconnection procedure. The developed solution, that is now presented in detail, is inspired by, and closely related to, the *K-method*; we however work directly and only on the discrete–time level without referring back to (possible) continuous–time origins of the discrete algorithms. I will finally apply the techniques inherited from the *K-method*, that are not explained in detail again; at the point I refer to [12].

In discrete time, the most general *resonator* consists of a discrete–time state

[28] A model of friction interaction has been implemented, that makes use of the structure and *resonators* presented here, and, indeed, a friction *interactor* with internal memory.

vector $\mathbf{w}$ and some "time-step" or "update" function R such that

$$\mathbf{w}(n) = R(\mathbf{w}(n-1), \mathbf{e}(n), \mathbf{f}(n)) \ , \tag{2.12}$$

where $\mathbf{f}$ is the output ("force") vector of the *interactor* (see figure 2.3.3) and the vector $\mathbf{e}$ represents some external influence on the *resonator*, that is *independent from* the *interactor*. [29] In plain words, with each time step, the *resonator* state is updated from the previous state vector and the contemporaneous external input vectors $\mathbf{f}$ and $\mathbf{e}$, coming from the *interactor* resp. some independent external source. Further on, the *resonator* shows a representing configuration vector $\mathbf{x}(n)$ to the "outside world", on which in turn $\mathbf{f}(n)$ depends. In this application here, a vibrating solid object is accessed through its "configuration", position and velocity, in a certain contact point (or area). $\mathbf{x}(n)$ "represents" (to the outside, in particular the interactor) the resonator in its state $\mathbf{w}$ via some function S:

$$\mathbf{x}(n) = S(\mathbf{w}(n)) \ . \tag{2.13}$$

Combining equations (2.12) and (2.13), we can see $\mathbf{x}(n)$ as a function of $\mathbf{f}(n)$ and the vectors $\mathbf{w}(n-1)$ and $\mathbf{e}(n)$, that are known from the previous time step resp. an external input:

$$\mathbf{x}(n) = S(R(\mathbf{w}(n-1), \mathbf{e}(n), \mathbf{f}(n))) \ . \tag{2.14}$$

The condition that is now imposed on the *resonator*, is for the concatenation $S \circ R$ of the functions R and S to split into two summands, one of which depends only on the known vectors $\mathbf{w}(n-1)$ and $\mathbf{e}(n)$, and another depending *linearly* only on $\mathbf{f}(n)$:

$$\begin{aligned} (S \circ R)(\mathbf{w}(n-1), \mathbf{e}(n), \mathbf{f}(n))) &= S(R(\mathbf{w}(n-1), \mathbf{e}(n), \mathbf{f}(n))) \\ &\stackrel{!}{=} T(\mathbf{w}(n-1), \mathbf{e}(n)) + L(\mathbf{f}(n)), \\ &\qquad L \text{ linear.} \end{aligned} \tag{2.15}$$

This condition is fulfilled in particular if both R and S are linear, as in our case of modal description with "pick up"points, or if both functions split in the described way. It is however thinkable that the condition holds neither for R nor S, but for the concatenation $S \circ R$, i.e. that non-linearities "cancel out". L as a linear mapping between finite-dimensional vectors can also be seen as a matrix $\mathcal{L}$ whose dimensions are the dimensions of $\mathbf{f}$ resp. $\mathbf{x}$ and we may write $\mathcal{L} \cdot \mathbf{f}$ instead of $L(\mathbf{f})$. If we now define $\mathbf{p}(n) \triangleq T(\mathbf{w}(n-1), \mathbf{e}(n))$, combine equations (2.16) and (2.14) to

$$\mathbf{x}(n) = \mathbf{p}(n) + \mathcal{L} \cdot \mathbf{f}(n) \tag{2.16}$$

and recall the definition of the *interactor*

$$\mathbf{f}(n) \triangleq F(\mathbf{x}(n)) \ , \tag{2.17}$$

[29] For clarity of the picture, $\mathbf{e}$ is not depicted in figure 2.3.3 as not relevant for the general idea and unproblematic.

we finally receive the crucial equation that determines $\mathbf{f}(n)$:

$$\mathbf{f}(n) \stackrel{!}{=} F(\mathbf{p}(n) + \mathcal{L} \cdot \mathbf{f}(n)) \; . \tag{2.18}$$

It is underlined again, that here $\mathbf{p}(n)$ does not depend on $\mathbf{f}(n)$, i.e. can be computed before $\mathbf{f}(n)$, and an implicit relation $\mathbf{p}(n) \mapsto \mathbf{f}(n)$ has been found, that completely coincides with the situation in [12], section C. This implicit relation (2.18) may be transformed into an explicit mapping — under the conditions of the *implicit mapping theorem* — or solved through an approximation; I refer to [12] for the detailed discussion that is not repeated here. It has to be noted from equation (2.16), that $\mathbf{p}(n)$ coincides with $\mathbf{x}(n)$ if $\mathbf{f}(n)$ is zero:

$$\text{If } \mathbf{f}(n) = 0, \text{ then } \Rightarrow \mathbf{x}(n) = \mathbf{p}(n) \tag{2.19}$$

In plain words, $\mathbf{p}(n)$ is equal to the output vector of the *resonator* under some fictitious "*pseudo-update*" with zero input (force). As a result, we finally see that the non-computable loop in figure 2.3.3, $\mathbf{f}(n) = F(?(\mathbf{f}(n)))$, can be turned into a resolvable implicit relation, equation (2.18), if the black box of the *resonator* is equipped with
1. a label containing $\mathcal{L}$ and
2. a *pseudo-update* functionality, that delivers the "simulated" *resonator* output with zero input, without de–facto updating the internal state.
The dimensions of $\mathcal{L}$ have already been mentioned as being of a similar order as those of $\mathbf{f}$ and $\mathbf{x}$; exactly, $\mathcal{L}$ contains $\dim(\mathbf{f}) \times \dim(\mathbf{x})$ elements. Passing $\mathcal{L}$ whenever necessary, i.e. when *resonator* or *interactor* or any of their attributes (such as modal parameters or the point of interaction for impact or friction) are exchanged, is thus a negligible overhead in comparison with the processing of the in- and output vectors $\mathbf{f}$ and $\mathbf{x}$ that have to be passed with each time step, i.e. usually 44100 times per second. In the concrete implementations here, $\mathbf{f}$ is one-dimensional and $\mathbf{x}$, and thus also $\mathcal{L}$, are two-dimensional. In particular is the size of $\mathcal{L}$ often small compared to the state vector of the *resonator*: the internal state vector of a *digital waveguide* e.g., can easily reach dimensions of the order of [30] 10000 while its representing external configuration would usually be of dimension 2 (position and velocity in a point. . .).

Summing up, the update-cycle at each time-step n for the complete discrete–time system consists of the following schedule:

In addition to the listed steps, at each change of a *resonator* or *interactor*, or any of their attributes, the values of the $\mathcal{L}$-*matrix* [31] have to be passed (possibly after recomputation).

Practical realization of the impact *modules*

The structures and algorithms described in the last sections have been implemented in *C* and combined into *modules* for the free [32] sound processing software

[30] A simple two-directional *waveguide* with a minimal frequency of 10 Hz at a sample-rate of 44100 Hz, contains at least two delay lines of 4410 samples each.
[31] The name has been chosen in analogy to the "*K-matrix*" of [12].
[32] Published under the *Gnu* open source license, as is the developed code.

pd [33]. *pd* executes *patches* of realtime audio processing and synthesis, consisting of interconnections of *modules*, which are atomic, separately programmed and compiled (usually in *C*) dsp blocks. The signal flow between such *modules* — internal ones, i.e. (usually standard) audio processing routines that are included components of the software, and *externals*, independently developed third–party components — is defined (in practice usually, but not necessarily only) by cabling boxes, representing *modules* or *subpatches*, in the simple graphical *pd* interface. The impact *modules* (and other developed audio processing algorithms) are, like all *externals*, linked into the *pd* environment in a plugin–like fashion, i.e. at runtime. While using only plain *C* (no *C++* code...), we generally apply an object–oriented(-like) programming style, as in the fragments of example-code shown below.

As a result of the modular architecture, the interconnection of *resonators* and *interactors* within the modules might also be accomplished at runtime. It is however not possible to define these connections within the *pd* environment, because *resonators* and *interactors* need to exchange information at least twice for every audio cycle (compare the update-schedule shown above), while *pd* processes the signal flow chunk-wise in audio buffers generally of a size of some hundred or thousand samples. [34] In the course of the *SOb* European project, the development of a "wrapper" module that could contain and manage, freely load and interconnect, *resonators* and *interactors* at runtime, rose as an intermediate

[33] Information about the principles, structure and handling of *pd* can be found in the various dedicated websites [55].

[34] It has to be noted that the size of the audio buffer in *pd* is specified in the program in *ms*. The standard value of 64ms thus corresponds to ca. $64 \times 44.1 \approx 2822$ samples. Even if this buffer is reduced to (practically probably problematic) value of one sample, mutual cross connection between *modules* would not be possible.

1. Read in external variables to the *resonator(s)*, such as additional external forces or related signals ($\mathbf{e}(n)$ in the notation above).
2. *Pseudo-update* of the *resonator(s)* from previous state $\mathbf{w}(n-1)$ and $\mathbf{e}(n)$, without de–facto update of the internal *resonator(s)* state. $\mathbf{p}(n)$ is passed to the *interactor*.
3. Calculation of $\mathbf{f}(n)$ from $\mathbf{p}(n)$. The mathematical technique for this step depends on the *interactor* function F. In the concrete cases here of impact an explicit formulation can be used in the (piece-wise) linear case, while the non-linear relation is solved through *Newton–Raphson* approximation [4].
4. After $\mathbf{f}(n)$ has been computed and passed to the *resonator(s)*, the internal *resonator* states are updated, $\mathbf{w}(n-1) \mapsto \mathbf{w}(n)$.

Figure 2.4: The update schedule at each time–step (sample cycle).

scope, but could finally not be accomplished within the temporal restrictions. The *resonator* and *interactor* algorithms are thus linked statically in the various *modules* described below. The original scope of the modular structure of the algorithms, however has been reached: discrete–time models of objects and processes of interaction have been and are developed independently, and combined later without the necessity to look (back) into, or adapt the internal structure. This reduces the costs and complexity of the development of the *modules* in various ways (e.g. through improved reuse of code by co-developers [35]) and forms a solid basis for the extension and functional enhancement (such as runtime linking) of the catalog of sound models developed in the course of the *Sounding Object (SOb)* European project).

A general resonating object in the modal description, as introduced in section 2.3.2, is implemented as a discrete–time algorithm in the "*modal resonator*". The differential equations of the modes (2.3) are discretized through bilinear transform, leading to a linear update equation. In particular, together with the linear equation connected to the *pickup* point (2.6), the resulting concatenation splits in the way described above (2.16) and the $\mathcal{L}$*-matrix* can be calculated from the modal parameters. Each mode appears as a second-order difference equation, or linear filter. The exact formulation of the discretization process, the resulting update equations and some considerations concerning the choice of using the bilinear transform are found in [58] and not repeated here. Figure 2.5 shows the first header lines of the *modal resonator* object, with the array of modes, each characterized by its frequency, decay time and weighting factors at the chosen *pickup points*. The $\mathcal{L}$*-matrix* is seen as the other *public attribute* of the object.

In many scenarios of contact, the inner vibrational movement of one of the involved objects can be neglected from the auditory standpoint because it is of very small amplitude compared to that of the other one. For a ping pong ball bouncing on the floor e.g., the vibration of the floor itself is hardly perceivable acoustically compared to that of the ball; or vice versa, a glass marble falling on a desk is hardly heard itself, since the vibration of the table caused by each impact is acoustically highly dominating. For the modeling of situations like these, it is (especially from a *cartoonification* standpoint) often sufficient to look at one of the contacting objects as a point-mass. An "*inertial resonator*" [36] has thus been implemented, that is very "cheap" and uncomplicated in its implementation. Notably, a free point-mass could also be characterized through its modal description of one mode with frequency 0 and no damping, i.e. infinite decay time in its impulse response. (The equation of a free point mass, $\ddot{\mathbf{x}} = \mathbf{f}/m$, coincides with (2.3) for $k = r = 0$.) The *inertial resonator* has been implemented explicitely in order to save some unnecessary computational overhead connected to the parameters set to 0, but mainly in order to simplify the control access: in the *modules* using the *inertial resonator* at the place of the general *modal resonator* the one control parameter of the mass replaces the parameters (resp.

[35] . . . as for the *modules* of friction already mentioned,

[36] This name aims at the fact that the mass, in its inertia — gravity does not play a role at this level (it does of course in higher–level models) — is the only attribute of this *resonator*.

inlets) for **1**. the number of modes (that would be set to 1, **2**. the number of *pickup points* (1 as well for a point mass...), **3**. the list of modal parameters and **4**. the list of weighting factors at *pickup points* (compare also figure 2.6) below).

Finally a *waveguide resonator* is being implemented, as an efficient alternative for objects with harmonic spectrum, such as strings or tubes. This *resonator* can be useful e.g. for the modeling of musical instruments or abstract structures, rather than everyday scenarios; it is not used in any of the *modules* or higher–level models presented in this thesis and thus not discussed here.

The above resonators can be simply used as linear filters (as has been done before in the case of the *modal resonator*...); in fact, the models also allow the direct input of external force signals to the *resonators* which makes their use as filters straightforward. Central point however is the mutual interaction of

```
typedef struct
{
    /**
     * Container for parameters of each mode
     */
    struct _modalobjb_modepubl
    {
      t_float freq0;
      t_float t_e;
      t_float *pick_contrib; /**< Array of weights of the mode
                                    at the interaction points */
    } *mode;

    t_matrix **pp_L;         /**< The L-matrix of to the modal
                                    resonator at the chosen/de-
                                    fined interaction point   */
  } publ;

  /**
   * Private parameters of the modal object
   * These are computed from public parameters by the function
   *                                set_privateprops_modalobjb
   * and should never be touched explicitly
   */
  struct _modalobjb_priv
  {
    /**
     * Container of filter coefficients of a mode
     */
```

Figure 2.5: The first lines of the code that defines the structure of the *modal resonator* with public and private properties of the object.

resonating objects as extensively discussed and described in the last sections. *Resonators* are connected in the style of figure 2.4, via an *interactor* describing the force occurring during impact interaction. For the interaction F in the 3. step of the update cycle (figure 2.4) the two alternative force terms of equations (2.1) and (2.2) have been implemented.

In the first, non-linear case (2.1), the characteristic relation of equation (2.18) can not simply be solved explicitely by an appropriate conversion. I apply the solution presented in [4] at the example of a one–mode (damped harmonic oscillator) *resonator*, and solve (2.18) through approximation via a *Newton-Raphson* algorithm. Details of the resulting computation and computational load can be found in [4] and [58]. Another possible method of solving this equation, as mentioned in the original derivation of the *K-method*, would be to store in a table the implicit function $p \mapsto f$ expressed through (2.18) via the *implicit function theorem*, and thus solve (2.18) at each sample cycle by a table lookup. This approach would be cheap in terms of computation during the realtime processing, but very problematic for control (in fact hardly usable in realtime), because at each change of any of the parameters reflected in the $\mathcal{L}$*-matrix* the lookup table would need to be recalculated and -filled.

In the case of the piece-wise linear impact force, equation (2.18) can be resolved directly, i.e. f can be isolated; the linear *interactor* is thus slightly cheaper in terms of computation.

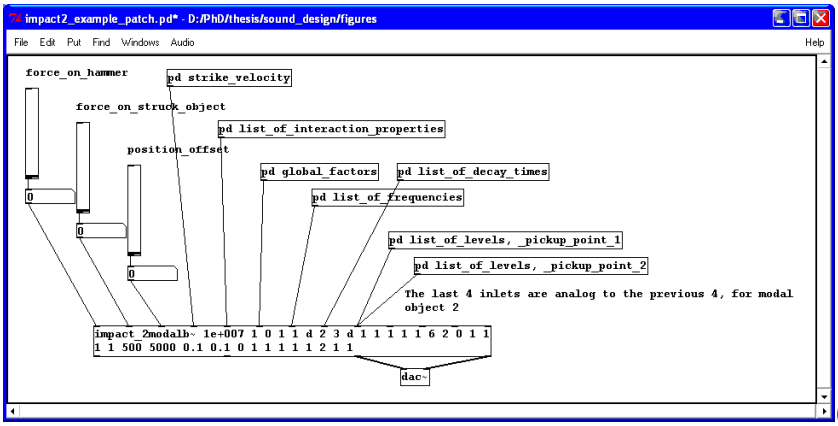
The presented *interactors* and *resonators* are combined in different *modules*. Four *pd* impact *modules* with *modal/inertial resonators* have have been implemented,
"`impact_modalb~`", "`impact_2modalb~`", "`linpact_modalb~`" and
"`linpact_2modalb~`", where the names reflect the used components. "`impact_`" refers to the non-linear, "`linpact_`" to the linear impact *interactor*. "`2modalb`" indicates the use of two modal *resonators* while the *modules* "`..._modalb`" use one modal and one inertial resonator. The final "`b`" represents the underlying method of discretization, *bilinear transform*; the realization of another modal *resonator* through application of a different method of discretization to the continuous–time modal description, may be a possible future task (that is why I chose to mark in the name the *bilinear transform* used here). Figure 2.6 shows the appearance of the *modules* "`impact_2modalb~`" and
"`impact_modalb~`" in the *pd*-GUI (i.e. their representing boxes), with according control connections. The latter one is used in all the higher–level models in section 2.4 as the somewhat best tradeoff in terms of computation, control and auditory potential.

2.3.4 General properties of the low–level impact model

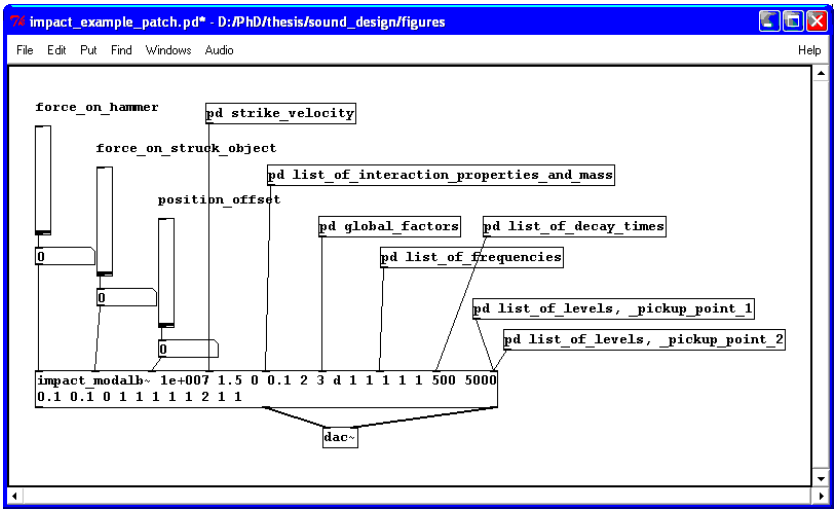
The general strengths of physics-based sound generation have been discussed in the introduction of this chapter, and we shall shortly look at some concrete consequences in the case of the impact model. I have mentioned the crucial significance of transient stages in contact sounds and the problematic of their insufficient description by existing signal-based theories. As a consequence of

its physical basis, the algorithm used here produces convincing and, if intended, realistic, transients that dynamically reflect all the involved physical control parameters. The impact parameters [37] with the strongest influence on the auditory appearance of the output of the algorithm, are the "hardness" k (see equations (2.1) and (2.2)) and the relation of the masses of the two objects.

[37] The complex of setting of parameters of the modal *resonator* is handled in the next section (2.3.5).



(a)



(b)

Figure 2.6: Screenshot of the *modules* "impact_2modalb~" (a) and "impact_modalb~" (b) in the *pd*-GUI. The modal parameters (frequencies, decay factors, levels — last 4 inlets in (a)) of the second modal *resonator* are replaced by just one mass parameter for the simpler *module* "impact_modalb~". This mass parameter is for practical reasons included into the list of "interaction parameters", although it is logically separate.

Exactly, the latter is not a parameter of the *interactor* but can still be seen as an impact parameter, since for one *resonator* used independently, without any mutual interaction, i.e. as a feedforward, linear filter, the mass only plays the role of an amplitude scaling factor. Again, to be exact, in the case of the modal *resonator* I should speak of impedance rather than mass; for the point-mass the two terms are equivalent (one being the inverse of the other) while for a distributed object e.g. the position at which it is struck is also significant for its external appearance in interaction, as the overall mass (or mass density). Important for the interaction however, is finally the relation of the impedances of the two objects (at the point of their interaction); I thus here vary only one of the two impedance values. Since the relevant effects are demonstrated in the following at the *module* with one modal and one simple inertial *resonator*, `impact_modalb~`, (that is also at the center of the following higher–level models) the mass parameter of the *inertial resonator* is used to control the "mass" relation (and therefore stick with this simplified term).

Figure 2.7 shows the trajectories of two contacting objects as modeled with the *module* `impact_modalb~`. [38] An inertial mass ("object 1") hits a *modal resonator* of three modes; shown is the position over time of object 1 and of the contact point of object 2 (in the direction of the one dimension of the impact model). Depicted below is the according distance value, with distance 0 corresponding to the lower boundary line of the window and "flipped orientation" (distance = trajectory 1 − trajectory 2), i.e. positive distances below the boundary line. Contact between the objects occurs where trajectory 1 is above trajectory 2 and the distance curves lies above the boundary line. The occurring interaction force is related to the distance via equation (2.1), 0 when the two objects are not in contact. In the examples of figure 2.7 λ, the dispersion, is 0, and the contact force is proportional to a "distorted" ($\alpha = 1.5$) version of the fraction of the distance curve inside the boundary lines. Without further detailed analysis, it is seen that such force trajectories are quite different from a semi-cycle of a cosine curve, as has been assumed and used in earlier works of synthesis of contact sounds [70][71], or other obvious simple profiles (such as an impulse...). For the two harder contacts, several "*micro-impacts*" occur, i.e. the objects contact and part several times until they finally stay separated. The examples are in accordance with the often assumed general rule that for a harder striking object, the higher(–frequency) modes of the struck object are excited increasingly. On the other hand it has to be noted that the trajectories during the phase of contact (i.e. until the last *micro-impact*) do not only consist of the components of the modes of one or both interacting objects. In fact, due to the complex non-linear interaction, it is a priori not clear at all, what a frequency domain representation of the transients during the contact phase would look like; further on it is basically unknown how such a representation would relate to auditory perception for such short signals. This points

[38] The figure 2.7 (as well as 2.8, 2.9 and 2.10) is a screenshot of the realtime output signal displayed in *pd*: unfortunately the labels ("object1"...) are not readable at this scale and axis ticks are missing; exact quantities however are not essential for the arguments given in the following, so that I avoid the effort reconstructing these examples.

out exactly the problematic of generating expressive contact transients with signal-based methods. It is not part nor a scope of this thesis to solve these obviously all but trivial questions; instead the model-based approach used here allows to efficiently exploit perceptual mechanisms without depending purely on signal-theoretic foundations.

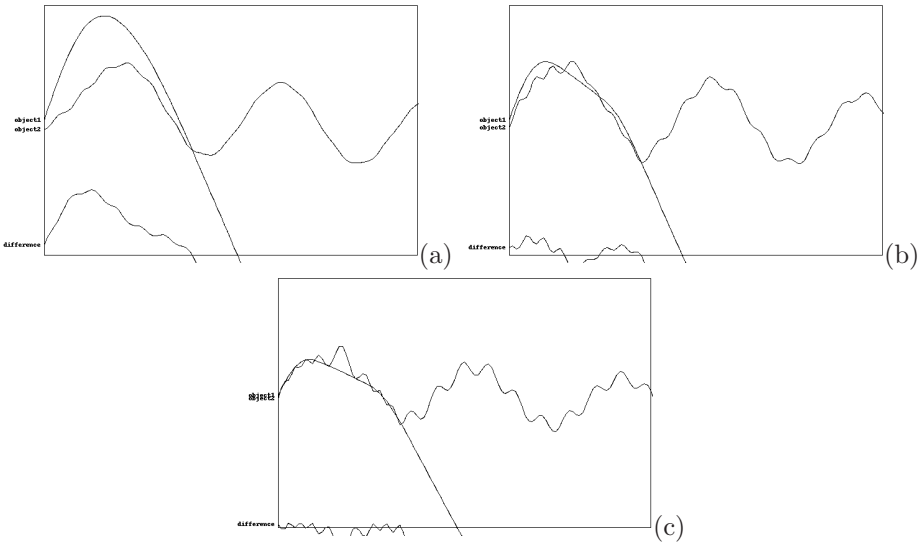


Figure 2.7: Inertial mass (object 1) hitting a resonator with three resonant modes/frequencies (object 2); the objects are in contact, when trajectory 1 is above trajectory 2, the distance variable is depicted below (shifted to the boundary of the window). The elasticity constant k, i.e. the hardness of the contact(surface) is increased from (a) to (c); note that for harder collisions several "micro-impacts" occur.

Figures 2.8 and 2.9 depict the influence of different mass relations on the impact signal. Again, it is seen that the distance curves, and thus the effective force profiles, are not easily described in terms of conventional elementary signals (such as sinusoids or impulses). In figure 2.8 it can be seen that only for very low relative mass of the striking object (1, relative to the mass/impedance of the modal *resonator*), the impact signal approaches an impulse response. [39] Thus, only for this limit case does it appear suitable to model contact sounds with impulse responses or filtered very short noise burst (as done before, compare [71]). Signals as in figure 2.9 may possibly rarely be found in mechanical "reality"; the perfect fitting of the modal description for a mechanical object over such a wide range of deformation should be the exception — mechanical objects break or undergo lasting deformation for applied forces above a certain level. (It also has to be kept in mind that here the vibration of objects at

[39] In figure 2.8(c) the phase of contact of both objects is so short that it is not visible in the display, i.e. in the order of 1ms...

one *pickup* point is modeled. In our surroundings, resulting acoustic signals that arrive at out ears or e.g. at a microphone necessarily look very different because of the spatial propagation of vibration from an area of the object through the air.) It is interesting that the depicted signals still sound "convincing", i.e. not completely unfamiliar or artificial: they still fit well the intended expression of a very heavy, "stiff" mass hitting a very "compliable" object. Interesting to note is the strong low–frequency impulse perceivable in the example of figure 2.9(b), while in the following decay the low frequency mode is hardly present. This somewhat contradicts the overall visible tendency of relatively strong low–frequency components for impacts with high relative masses and vice versa comparatively dominant high frequency parts for very low masses, as notable in figure 2.8. Generally, the signal-theoretic (Fourier-based) description of the shape of the signals in figures 2.8 and 2.9 during contact, is again not

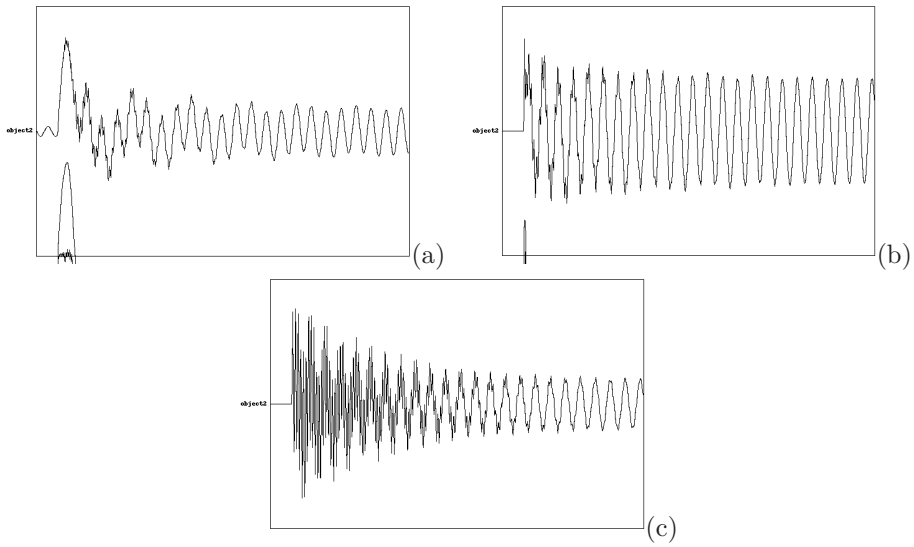


Figure 2.8: Impacts with decreasing hammer mass (from (a) to (c)).

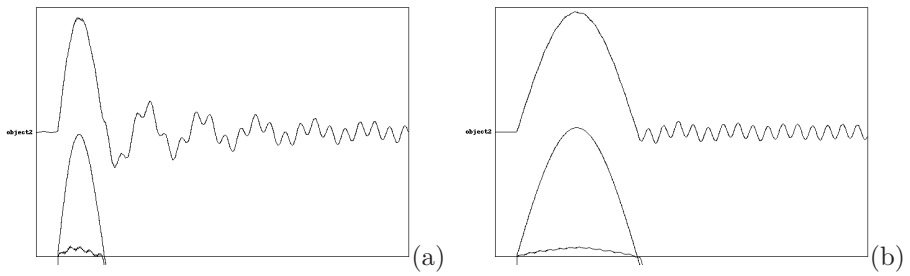


Figure 2.9: Very high relative hammer masses. For many mechanical resonators these examples should exceed the usability of the linear description.

satisfying or particularly helpful. Again, I do not intend to give, nor search for, any deeper explanations about the perception of the presented phenomena. A solid physical, mathematical or psychoacoustic analysis of transient sounds is not the scope nor the field of this work. On the contrary, strengths of a physics-based approach are shown for sound modeling, beyond the restrictions of current psychoacoustic knowledge.

The finally, for the higher–level models (presented in the following section) maybe most important characteristic of the impact algorithm, as compared to sample-based sound, is its dynamical behavior (mentioned already in the introductory lines). Figure 2.10 shows two generated trajectories whose difference is only due to different initial states of the modal object before contact. All other parameters, including the initial hammer velocity are equal in (a) and (b). It is seen that the profiles of interaction as well as the following decay stages can vary remarkably. This is in strong contrast to the static nature of repeated samples and very important in cases of frequent or continuous contact. In particular, the model of rolling presented later would be impossible to realize on the basis of fixed, prerecorded/stored impact components.

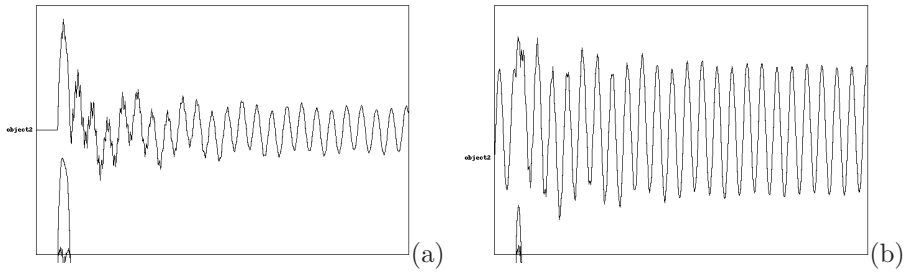


Figure 2.10: Impacts with identical parameters as in figure 2.7 (c) on a larger scale; the "hammer" trajectory is shifted as the distance trajectory for a clearer view. The difference between (a) and (b) is only due to the different states of object 2 at first contact.

2.3.5 *Resonator* attributes

I have listed the advantageous properties, that led to the adoption of *modal synthesis* for the lowest–level physics-based models, and mentioned different possible strategies for the tuning of modal parameters (in section 2.3.2). These remarks are now concretized and applied.

For struck solid objects, the distribution of the modal frequencies and decay times, or equivalently: widths of resonant peaks, generally depends on the material and shape of the object. In the case of special constraints, such as a tautened drum skin, further parameters, such as the stress of the membrane, can be important. The weighting of the modes, i.e. the individual level of excitation in an impulse response, further on depends on the positions of in- and output, i.e. of the attacking force and the measured response. When using the modal

formalism for the modeling of contact sounds, the choice of possible or suitable strategies for the tuning of modal parameters, depends on the scenario to be modeled and the modeling paradigm (such as simulation or abstraction). For some rare examples, such as idealized circular or square membranes or clamped bars, modal parameters can be found analytically, directly from solutions of the differential equations that describe the system. In most cases however, analytical solutions are not known, and the modal parameters must be found or chosen in a different way.

When modeling one specific existing mechanical object, the modal parameters can be approximated from measurements or recordings of the movement of the object. Ideally, special excitation and pickup devices are used, that allow to induce an exactly chosen force, e.g. of the form of a sinusoid, impulse or noise, at one point and to measure the movement of the object at another (or the same) point, with no (e.g. optically...) or negligible interference. Recorded impulse responses can be accurately decomposed into components of exponentially decaying sinusoids, by *High Resolution (HR) Analysis*, a dedicated analysis method (e.g. [37]). Van den Doel et al. have examined in depth auditory effects of simplifications in the modal description of synthesized sounds [72] and demonstrated a rather large robustness of basic characteristics. This potential of extensive but well-directed simplifications is the basis of the use of modal parameters in the following work. We have seen in section 2.3.2 that each mode appears as a resonant lowpass filter with a peak near its frequency, and that the frequency response of the whole object is a weighted sum of its "mode filters". From this observation, for rougher, more cartoon-like, modeling, the frequencies of the most prominent modes can also be read approximately from a frequency response, whether directly measured (e.g. through sinusoidal input forces) or as a frequency representation derived as Fourier transform of a recorded time-scale signal. Even in non-ideal recording conditions, the prominent modes can be identified from peaks in the response. This last approach has been used with recordings (with a standard microphone) of a "*Bodhran*", an Irish frame drum, struck at different points, to tune the impact *module* as a simple cartoon model of the instrument allowing for dynamic control inspired by the "real" object. Details and connected control interfaces are described in the following chapter (section 3.1).

Another way of deriving the modal parameters of a specific object, is to first construct a highly exact (and accordingly complex and computationally expensive) finite–element description on the basis of the exact specification of the shape and material properties, and to extract the modal parameters from this computational description in an analog way as from mechanical measurements. This technique is used by the dedicated software "*modalys*". The general direction under the present premises of ecological expression (rather than simulation) is somewhat opposite. Central starting point is the question of what ecological attributes are, or can be, perceived from the sound of contacting objects. The next consequent step is then to ask how such attributes (considered worthwhile) can be expressed through the models. Fontana et al. have conducted experiments addressing the perception of basic shapes (such as spheres and cubes)

from simplified sounds of hollow cavities [26]. The modal *resonator* has been tuned according to parameters used here, with an interpolation mechanism that allows to morph between the proposed characteristics of spheres and cubes; this patch has in turn been used in listening experiments.

While the human capability of auditory recognition of shape (per se, without previous dedicated training) may still offer wide space for questions, the auditory convection of material properties has been recognized of clear potential [40]. In fact, the categorization/recognition of material from the sound of struck objects is everyday experience [40], and the inclusion of mechanisms of material expression into the modeling efforts reported so far is a promising (and significant also for following higher–level models). Several studies exist in the topic, focusing on different aspects, starting from different assumptions and following various strategies, with consequently different results. Lutfi and Oh [46] examine material perception from the resonance behavior, more exactly impulse responses, of "ideal" bars of fixed shape, ignoring internal friction. Klatzky, Pai and Krotkov [40] 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 [73] use the latter experiences to lay out a broader method of rendering sounds of hit objects, under additional consideration of interaction position. Avanzini and Rocchesso have used [3] a preliminary version of the presented impact *module* (with a damped–harmonic oscillator, a *resonator* of one mode) in a listening-test of material expression. They have found that even with one resonant mode a rough material categorization can be achieved (in a forced–choice test) with influences of both mode frequency and decay time. In the present context of rather abstract modeling — aimed at here are generally no highly concrete [41] but rather generic scenarios — the strategy of a proposed shape-independent material characteristic as in [40] is very well suited. This approach is based on a pioneering work by Wildes and Richards [78] that derive a material specific coefficient of internal damping (as an approximation from material properties). In the modal description, this damping coefficient ϕ appears as a slope factor, where the decay time of the modes is antiproportional to the modal frequencies. Some further details have been worked out and can be found in [40]. The strategy has been adopted in a patch where modal decay factors are calculated from mode frequencies following the material-characteristic damping coefficient. As in [40] another factor of "external damping" is included that represents the loss of vibrational energy due to friction e.g. in the surrounding air. The method, that has been supported through psychoacoustic testing [40][72] before, is here completed with the the physics-based model of the impact itself; the resulting capability to include further (material- /surface-specific) interaction parameters, such as hardness of contact, fundamentally contributes towards expressivity and realism. Of course these examinations would open up a wide

[40] Probably everybody can share the experience that a struck glass will just from its sound not be confused with a struck wooden object...

[41] The example of the *Bodhran* mentioned above is the only case where modeling efforts started from a distinct concrete object.

field for systematic testing. One should also keep in mind that the linear decay/frequency dependence is one possible approximation and psychoacoustic studies e.g. also show a slight influence of frequency ranges on material impression(compare e.g. [3]). Practical sound design examples can benefit from intuitive deviations of modal parameters from exact theory-based values. On the background of a missing overall closed theory of the auditory capabilities and mechanisms of material perception (while the general phenomenon of auditory material recognition or connotation can not be doubted in its existence and significance) the intuitive accessibility of model parameters may suggest 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, for different design goals can be a rewarding challenge. In the higher-level models, modal parameters are often tuned without "mechanically" following any strict existing formalization, but in awareness and use of discovered (in the works cited above) tendential connections of modal parameters and material impression.

Also the depiction of the position of interaction through the position dependent modal weights, can be approached in various ways. Again, weighting factors may in some cases be gained exactly after theoretical considerations, where an analytical solution of a specific system is known [42]. As generally mentioned above, alternatively, either accurate numerical simulations (e.g. finite–element methods) or "real" physical measurements can be used. For the cartoon model of the *Bodhran* e.g., the position depending weighting factors have been tuned, as all modal information (as already mentioned), after (microphone) recordings of the struck instrument at several points. (The dedicated section 3.1 in the following chapter gives more details.) From an even more abstract, *cartoonification* standpoint, qualitative observations on modal shapes (compare figure 2.2 as an example) are useful and important to note: for modes of higher frequencies the number of *nodes* increases and its spatial distance accordingly decreases.
1. One consequence is that for higher modes even small inaccuracies in interaction or pickup position may result in strongly different weighting factors, so that an element of randomization can here add "naturalness". [43]
2. For interaction positions close to a boundary, which is a common node for all modes, the lowest modes gradually disappear and higher modes (with smaller "regions of weight") relatively gain in importance. This phenomenon can be well noticed for a drum and is strongly present in the analyzed recordings of the *Bodhran*: if the membrane is struck close to the rim, the excited sound gets "sharper", as the energy distribution in the frequency spectrum gets shifted upwards ("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

[42] E.g., 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.

[43] Such a random contribution is used in the setting of modal gains in the model of a bouncing object as described in section 2.4.1.

modes, those with modal shapes antisymmetric to central axes, are not present in the center of a round or square membrane. They consequently disappear "bottom–up" when approaching the center point; again this notion has been supported and used in the analysis and modeling of the *Bodhran*.

Finally, the generally very clear perceptual meaning of the modal description (sinusoids with envelopes, resonant peaks...) always has to be kept in mind: in the present premises of sound design (as opposed to simulation), modal parameters can also sensitively be "tuned by ear", the ear being the final judging instance.

2.4 Higher–level scenarios and structures

2.4.1 Bouncing

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 the impact algorithms allows the input of an external force term. A bouncing process can be simply achieved with an additional constant term representing gravity. Figure 2.11 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.

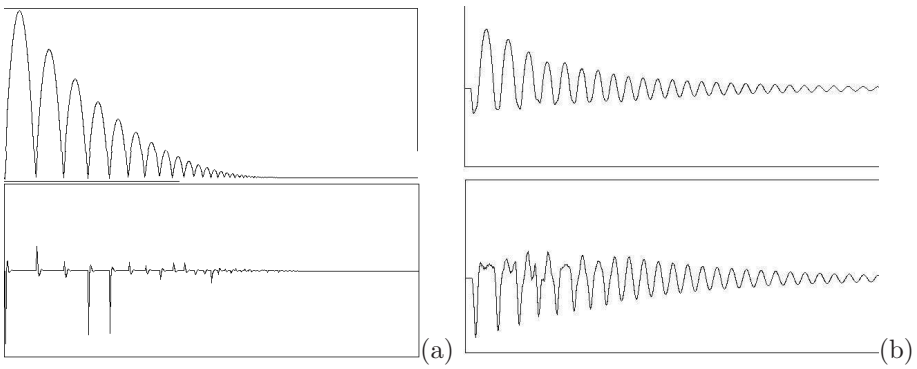


Figure 2.11: An inertial mass "bouncing" on a two-mode resonator. (b) focuses on the final state of the process: The two interacting objects finally stay in constant contact, a clear difference to simple repeated samples.

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 level of the audio algorithm 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 (three-dimensional) reality only for perfect spherical objects or special, highly restricted, symmetric situations. These restrictions led to the development of a "bouncer" control structure, that explicitely creates typical patterns of falling objects. Underlying considerations are sketched in the following.

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 about 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 negligence of friction in surrounding air, 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.

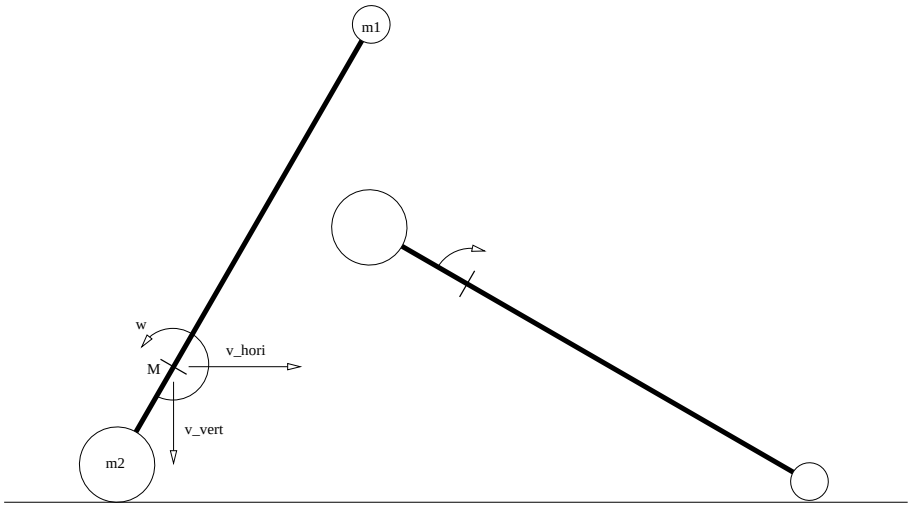


Figure 2.12: A non-spherical object bounced at the ground in two different states. Here, a particularly clear example is chosen, a "stick" with its mass lumped at both ends. The rotation is in both cases about an axis parallel to the ground.

We start with the analysis of the free movement of a bouncing object that is bounced at the ground at time $t = 0$ with an upward vertical velocity $v(0) = v_0$ of its center of mass. For a constant gravity acceleration g, v decreases according

to

$$v(t) = v_0 - g \cdot t, \quad g > 0, \tag{2.20}$$

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

$$x(t) = v_0 \cdot t - \frac{g}{2} \cdot t^2. \tag{2.21}$$

During this free rebounce 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 retransferred to E_{kin}. Both terms reach their 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

$$t_{return} = \frac{2}{g} \cdot v_0, \tag{2.22}$$

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 friction forces in surrounding gas are neglected, 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 and velocities here coincide with their respective vertical components. The amount of energy "loss" with reflection is exactly generally different for each impact, as can be seen e.g. from figure 2.10 in section 2.3.1, 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 before and after, E_{pre} and E_{post}, each reflection,

$$E_{post} = C \cdot E_{pre}, \; C < 1, \tag{2.23}$$

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

$$E(n) = C^n \cdot E_0, \quad v(n) = \sqrt{C}^n \cdot v_0, \quad t_{int}(n) = \sqrt{C}^n \cdot t_{int}(0). \tag{2.24}$$

The implementation of this basic scheme in fact delivered very convincing results in comparison to the afore-described implicit pattern simulation. In figure 2.11 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; in scenarios, labeled "bouncing" though, 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 2.12). 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 realtime 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 both, if friction forces are neglected as in the model of impact interaction, 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, it can be said 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 important impact position, vary due to rotation. Consideration of these deviations, especially 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, also of non-spherical, even asymmetric, objects, occurring 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.

Summing up these observations, 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/-velocity, together with
2. the initial impact velocity.
3. The acceleration factor is the quotient of two following maximal "bounce-intervals" and describes the amount of microscopic energy loss/transfer with each reflection, thus the speed of the exponential time sequence.
4. The velocity factor is defined analogously.

   Note that these parameters should for a spherical object be equal (see above), 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.
5. Two parameters specify the range of random deviation below the (exponentially decaying) maxima for temporal intervals resp. impact velocities. 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.

2.4.2 Breaking

The auditory perception of breaking and bouncing events is examined in a study by Warren and Verbrugge [77]. It is shown, that sound artefacts, 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 subevents of

emitted, falling and rebouncing fragments. Some further thoughts strongly help on successful modeling: Typical fragments of rupture are of highly irregular form and 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 "notion". Supporting Warren and Verbrugge's examination, a short noise impulse added to the attack portion of the pattern underlined the breaking character.

As another insight during the modeling process, several sound attributes showed to be important. Temporally identically grouped impacts 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 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 2.3.1, e.g. figure 2.8, 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.

2.4.3 Rolling

Particularly rich in ecological information are the sounds of rolling-scenarios: 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* [30] attributes, such as velocity, gravity or acceleration/deceleration. A series of dedicated psychoacoustic studies [35, 36] has been dealt with these perceptual phenomena. This suggest acoustic modeling of Rolling to be a rewarding goal under the various demands of auditory display.

In fact, the value of rolling-sounds has been recognized before and resulted in sound synthesis works [68][70]; but here, a simple source–filter approach shows to be of restricted applicability and according sound results are only partly convincing. Assuming fixed force profiles, the distinction of rolling and sliding

e.g. is not a priori clear, which is reflected in received sonic results. Everyday experience on the other hand tells, that the sound produced by a rolling object is usually recognizable as such, and in general clearly distinct from sounds of slipping-, sliding- or scratching-interactions, even of the same objects. The specific dynamics of rolling-interaction is not sufficiently captured just by low-pass filtering of surface profiles, as will also be substantiated from a geometrical viewpoint below (section 2.4.3). A physics-based approach is therefore applied, keeping in mind the general premises (*cartoonification* a.o.) as fixed in section 2.1, and avoiding the complexity and computational overkill of a complete three-dimensional, e.g. finite–element, simulation.

Its mentioned distinctive auditory character maybe partly seen as a consequence of the nature of rolling as *the continuous* interaction process, where the mutual force on the involved objects is described as an impact without additional perpendicular friction forces: in contrast to rubbing-, sliding- or scratching-actions, additional forces parallel to the surface are very small. [44] In most contact scenarios based on microscopic impact, on the other hand, phases of continuous contact (i.e. where single micro-impacts are not clearly distinct) are rather rare and insignificant. In bouncing processes e.g., continuous contact in this sense occurs only shortly before the objects come to rest and vibration is thus very small, usually hardly acoustically perceivable. Modeling bouncing-scenarios, I have noted the lower significance of these details and consequently omitted them in the *cartoonification* process. The situation of rolling is somewhat contrary: the interaction of the two involved objects (the rolling one and the plane to roll on) basically stays in this phase of continuous contact and distinct bounces occur only occasionally (e.g. caused by larger irregularities in the objects, especially at higher velocities). In simple words, rolling can be seen as bouncing on a smaller scale, and vice versa. Such characteristic details are exploited and magnified in the sound design concept applied here. The global movement of the rolling object does not need to be included and simulated in complete detail to account for the main auditory cues. Instead of expanding the closely physics-based impact model to the three-dimensional rolling-scenario, the global geometry is "reduced" to the one dimension of the efficient one-dimensional algorithm. The most important macroscopic features of the scenario are later accounted for explicitely. The development of an expressive real-time sound model of rolling in the hybrid hierarchical architecture is described in the following.

Reduction of local rolling-geometries to one (impact-) dimension

The acoustic vibration in a rolling-scenario has its cause in the structures of the contacting surfaces; no sound would emerge if the rolling object and the plane (on which it is rolling) had perfectly smooth surfaces — or at least, no other than a possible limited decaying vibration as for a vertically falling object. In fact, as an object rolls, the point of contact moves along it's surface and along the

[44] Probably the main notion behind the invention of the wheel...

plane. These "tracked" surface profiles are the source of the acoustic vibration in rolling-interaction.

If we restrict our view on the scenario to the one dimension perpendicular to the plane, the tracked surface profiles, exactly their difference, give rise to a time-varying distance-constraint on the interacting objects (i.e. the rolling object and the plane). This constraint takes the form of a temporarily changing distance-offset that adds to the distance variable x in equation (2.1) as it would emerge from the movement of the interacting objects. In other words, the surface profiles are the origin of a dynamic offset signal that has to be fed into the impact model, namely added to the distance-variable x, thus causing vibration of the contacting objects. Exact investigation however reveals, that

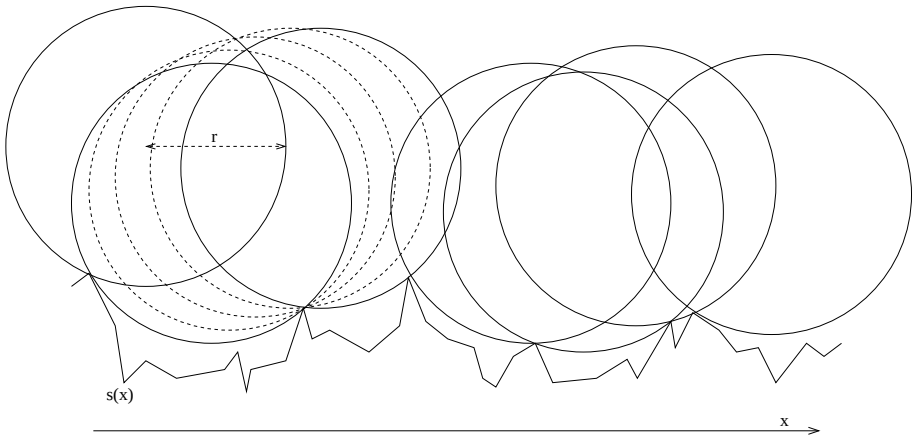


Figure 2.13: Sketch of the fictional movement of a ball, perfectly following a surface profile s(x). Relative dimensions are highly exaggerated for a clearer view. Note that this is *not* the de–facto movement; this idealization is used to derive the offset-curve to be used by the impact-model.

the appropriate offset signal is *not* simply the difference of the surface curves, as scanned along the rolling trajectory: not all these surface points (along the trajectories) are possible points of contact. Figure 2.13 shows the principle of rolling-typical "bridging" of surface details. The rolling object is here assumed to be locally perfectly spherical without microscopic 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 however showed to be satisfactory for the present modeling aims. It is seen that only certain surface "peaks" are potential contact points. The *hypothetical* trajectory of the rolling object, i.e. precisely its center, as depicted in figure 2.14, as it would move along the plane at constant distance 0 contacting the plane exactly at these peaks (without "bouncing back" or "enforced contact", i.e. distances ≤ 0, figure 2.13),

is finally the offset curve that expresses the constraint on the objects. *The actual*

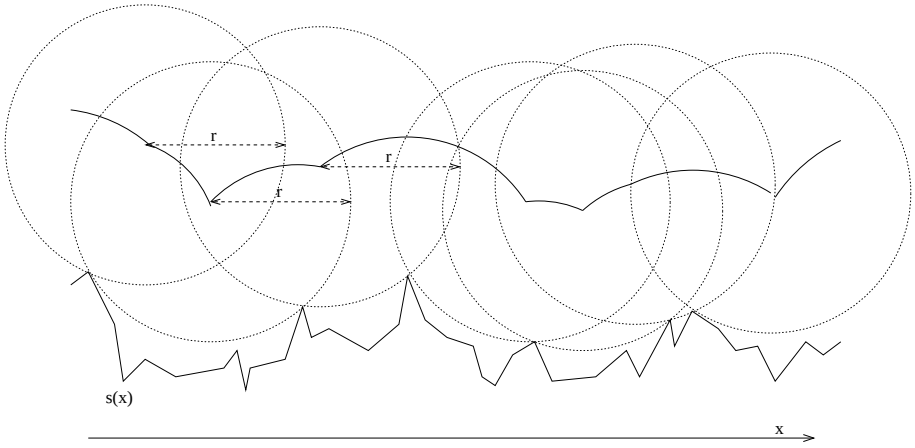


Figure 2.14: Sketch of the effective offset-curve, resulting from the surface $s(x)$. The condition on the surface to be expressible as a function of one curve parameter x is clearly unproblematic in a "rolling" scenario.

movement of the rolling object differs from this idealized trajectory due to inertia and elasticity. It is exactly the consequences of these physical properties, which are described by, and substantiate the use of, the impact model.

Implementation of the "rolling-filter"

In a straight approach, the calculation of contact points, necessary for the subsequent generation of the offset signal, is computationally highly demanding: in each point x along the surface curve, i.e. for each sample-point in a discrete implementation at audio rate, the following condition, which describes the momentary point of contact p_x, would need to be solved.

$$\begin{aligned} f_x(p_x) &\stackrel{!}{=} max_{q\in[x-r,x+r]} f_x(q) \qquad where \\ f_x(q) &\triangleq s(q) + \sqrt{r^2 - (q-x)^2}, \quad q \in [x-r, x+r] \end{aligned} \tag{2.25}$$

The ideal curve would then be calculated from these contact points. E.g. for a diameter of $10cm$, a transversal velocity of $1m/s$ and a spatial resolution according to an audio sampling rate of $44100Hz$ at this tempo [45] the above operations, maximum/comparisons and calculus, had to deal with $44100 * 0.1m/1m = 4410$ values at each sampled position, i.e. 44100-times per second. Of course these computational costs are high in a real-time context for standard hardware, especially in the context of sound cartoons to be used within wider (also multimodal) environments of human-computer interaction. The computations might

[45] ... i.e., if the surface profile is assumed to be resolved with a resolution such that when tracing the surface at the velocity of $1m/s$ samples appear at $44100Hz$, a canonical choice...

be executed offline, which would however restrict the realtime reactivity of the model; object radius and surface structure had to be fixed and could not be easily changed dynamically.

The solution comes in form of a recursive algorithm that solves the described task with a highly reduced number of operations, to the order of 10 in average per sample, and therefore minimizes the computational load enabling realtime implementation. Computational costs are here comparable to that of a lowpass filter or other simple approximations that have been developed and tried by the author (figure 2.15 sketches an example). In fact, lowpass filtering has been suggested and used to simulate the acoustic effect of rolling but sound results are quite different [70]. This is not surprising when remarking that the offset-curve as in figure 2.13 can contain strong high-frequency components (connected to its "edges"); such high frequencies may in some cases even be stronger than in the originating surface-profiles, contradicting the idea of lowpass filtering. Even the heuristic and computationally simple approximation sketched in figure 2.15 appeared comparatively more useful.

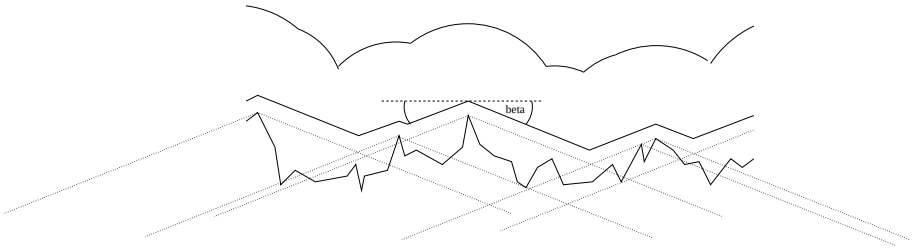


Figure 2.15: A simple approximation of *rolling-filtering* with the ideal offset curve above, for comparison. This trajectory would origin from "ideal" (frictionless, continuous contact) sliding of an angle along the surface. It can be noted that the detected points of contact are not exactly the same as in the idealized rolling of figures 2.13 and 2.14.

Surface

Different origins can be thought of, for the surface profile, which is a basis of the rolling-model developed above. 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 support well the preliminaries of our modeling efforts: expressive, flexible and effective sound cartoons are at the point of interest rather than fixed realistic simulations of single specific scenes. Stored sound/signal files are generally hard to adapt to varying model attributes.

The use statistics-based "surface"-models is thus preferable, that can efficiently generate signals of varying attributes. It is common use in computer graphics to describe surfaces in fractal parameters. One application of this idea

to the 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.

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 lost in significance. A very coarse approximation of this frequency curve was therefore chosen, 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, can/must also be carefully adjusted to surface (e.g. material properties) and strongly contribute to the expressiveness of the model.

Explicit modeling of macroscopic characteristics

Typical rolling-sounds usually show periodic patterns of timbre and volume that are of high perceptual importance. Periodicities that originate from macroscopic deviations of the rolling-shape from perfect sphericity — or more general, asymmetry of the object with respect to its center of mass — appear to form one important auditory cue for the recognition of rolling-sounds from similar sounds of contact, e.g. sliding. Also, the frequency of such periodic patterns strongly influences the perceived transversal velocity of the rolling object. Global asymmetries lead to modulations of the effective gravity force, that holds down the rolling object, an effect that gets stronger with increasing velocities (as motivated below). Usually less dominant is the simultaneous oscillation of the instantaneous velocity (of the point of contact along the plane). Such effects have to be explicitely accounted for by according parameter modulations, since the physics-based core is one-dimensional and does not cover higher macroscopic geometries.

Figure 2.16 sketches an asymmetric rolling object in different positions. Its center of mass is accordingly at different heights giving different terms of potential energy. In a free rolling-movement these oscillating terms of height of the center of mass $c(t)$ and potential energy are coupled to accordingly oscillating terms of kinetic energy and thus momentous velocity. This periodic energy transfer is connected to a periodic term of force acting between the rolling ob-

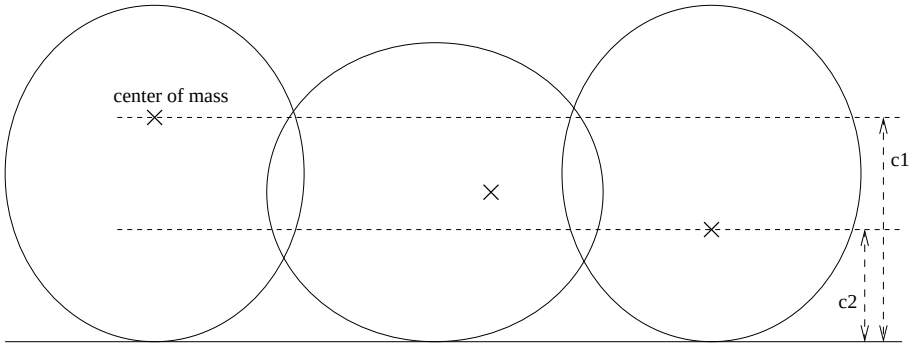


Figure 2.16: Sketch of a rolling object at different instants, (strongly) asymmetric with respect to its center of mass.

ject and the plane (in addition to the constant gravity force). The exact terms of forces and velocities effective in this (free rolling-) situation could be found as solutions of the differential equation given by stating the principle of energy conservation; they can of course only be determined if the shape of the object is known exactly. However, in the context of effective *cartoonification*, I derive a simple example-approximation in the following, that reflects the general behavior. (With our goal in mind, ecological expressiveness rather than simulation for its own sake, it has to be considered that the *exact* shape of a rolling object is rather not perceived from the emitted sound? A general idea of "asymmetry" however may be given acoustically.)

It is assumed that the oscillating (in the sketch of figure 2.16 between the extrema of $c1$ and $c2$) height of the center of mass $c(t)$ is approximately described by a sinusoid [46] .

$$c(t) = (c2 + c1)/2 + (c2 - c1)/2 \cdot \sin(\omega t) \tag{2.26}$$

The offset force-term between the two contacting objects (the rolling and the plane) is then connected to the acceleration perpendicular to the plane through Newton's law $F(t) = M \cdot \ddot{c}(t)$, where M is the overall mass of the rolling object. The acceleration is the second derivation of equation (2.26).

$$\ddot{c}(t) = -(c2 - c1)/2 \cdot \omega^2 \cdot \sin(\omega t) \tag{2.27}$$

This sinusoidal force modulation term proportional to the square of the velocity in fact gives convincing sound results despite all involved approximations; a constant modulation amplitude sounds unnatural for changing velocity. In the model, a parameter of asymmetry, in these equations $c1-c2$, allows to express an overall amount of deviation from perfect spherical symmetry. The modulation

[46] This is e.g. the case for a spherical object rolling with constant angular velocity (which may in free rolling be approximately the case for small asymmetry or a forced condition) whose center of mass is located outside the geometrical center.

frequency ω is related to the transversal velocity v and the (average) radius r of the rolling object, through $\omega = v/(2\pi \cdot r)$.

Chapter 3

Interaction examples

In the following, some example interfaces are presented or interactive devices that combine sound models developed in chapter 2 with gestural input and/or graphical display. The first two examples resulted from collaborative work, within the European project "*The Sounding Object* (*SOb*)" [67] and are rather subsidiary for the thesis as a whole, but serve to practically exemplify some of the potentials and strengths of the developed techniques and models; they are handled rather briefly, details about the various contributions of the involved collaborating institutes can be found in the dedicated publications that are cited in the according following sections. The last example device however, the *Ballancer*, is of further importance as it is used in the following chapter (4) for evaluation experiments to demonstrate the suitability and success of the sound design concept in reaching the initial scopes from which the work started.

3.1 The *Vodhran*

Behind the "Virtual Bodhran" or "*Vodhran*" stands the idea of a realtime cartoon model of a traditional Irish frame drum, the "Bodhran", in its playing style and sonic behavior. To this end, the `impact_modalb~` (section 2.3.3) *pd*-module is in its *resonator* properties tuned towards the behavior of the real instrument and connected to a realtime interface that allows drum–like playing control. Several alternative mechanical interfaces were used with individual advantages and disadvantages whose discussion is beyond the scope of this thesis; a detailed description can be found in [14]. The *Vodhran* is of interest here mainly to exemplify the idea of *cartoonification* applied to the sound emitting scenario, the played drum in its various aspects, and the robust practical handling of modal synthesis in a "pragmatic" approach.

The Bodhran is a frame drum, i.e. it has only a very small hollow resonant cavity and consists basically of a circular tautened membrane whose movement (after being struck) dominantly determines the sound of the instrument. Due to the very flat dimension of the frame, with typical depths of 4 to 9 centimeters in

relation to diameters of 30 to 50 centimeters, the membrane which is tautened on it can easily be accessed from both sides. It is usually struck with a typical "double–headed" stick with one hand and touched with the other hand from the opposite side to control an effect of damping, ranging from slightly "muffled" vibration to very strong "muting" with short decay. The sound of the membrane also varies characteristically depending on the position where it is struck, more exactly, how close to the edge (the frame) or center: in accordance with a general observation made about the modal description and its practical consequences in section 2.3.2, the higher–frequency modes gain in relative weight towards the edge (which forms a common node for all modes) while at the center over the whole frequency range certain modes disappear for reasons of symmetry. The cartoon model tries to account for (and possibly exaggerate) these prominent characteristics without necessarily simulating the complete instrument in all acoustic details as realistically as possible. With the very concrete scope of modeling the Bodhran — the individual differences between different exemplars of the instrument in construction and thus in sound are rather small (not like for some other instruments, e.g. guitars) — modal values were extracted from recordings of one drum struck at several positions between center and rim. As already stated in section 2.3.2 the frequency response of one mode of an object, here the Bodhran, is that of a resonant lowpass filter; the response of the whole object accordingly is a sum of such resonant filters, a parallel filter bank with peaks at the frequencies (exactly: near) of the most prominent modes. The appearance of the prominent modes in the frequency response is "robust" enough to allow their approximate extraction even on the basis of microphone-recorded signals of a struck object, despite all the involved inaccuracies: the force appearing in a stroke is not an impulse but approaches this theoretical profile for small masses (compare also section 2.3.4), striking interaction is spatially distributed but the contact area can be kept small (by using a small striker), and the wave distribution through air does not blur the main peaks if distance and reverberation are limited sufficiently. In this way, the frequencies of the 18 most prominent modes of an example Bodhran were extracted from recordings of the instrument struck at 5 equidistant points between the center and the edge. These frequencies and the according decay times, that were calculated from spectra at the beginning of the decaying sound and after a fixed time, are independent of the position of the stroke. The relative weights, i.e. levels, of the modes depend on the point of interaction. In the final model the modal weights are adjusted according to the position of the virtual stroke, interpolating between the 5 skeleton values. The damping of the membrane (usually with the left hand) is accounted for in a very *cartoonified* manner by simple proportional shortening of the decay times of all modes. Together with the various control interfaces (see e.g. figure 3.1) that are not discussed here (compare [14]) virtual drum instruments, "cartoon versions" of a Bodhran are reached that can be played in a similar way as the real instrument in its main features.

3.2 The *Invisiball*

Our [1] first try to add a tangible control interface to the sound model of rolling was based on the idea of a ball rolling on a deformable surface. In its "relaxed" resting position the virtual surface is perfectly plane and horizontal, so that the (virtual) ball without external interference (i.e. deformation of the surface) keeps rolling straight in one direction, slowing down only due to friction (in the movement). The virtual surface can be "bent in", given a "dell", by pushing on its mechanical representation, an elastic cloth tautened on a rectangular frame (see figure 3.2). In this way, the virtual ball receives an acceleration towards the center of the "dell", corresponding to the point where the cloth is pushed down, due to the occurring slope of the surface and gravity. The depth of the surface profile, and thus the strength of the resulting acceleration are proportional to how far the representing elastic cloth is pushed down. Position and depth of the control (pushing) movement are measured with the *Radiobaton* [56], a position controller developed by Max Mathews. The *Radio-Baton*

[1] The *Invisiball* as the *Vodhran* emerged from collaborations within the *SOb* project [67].

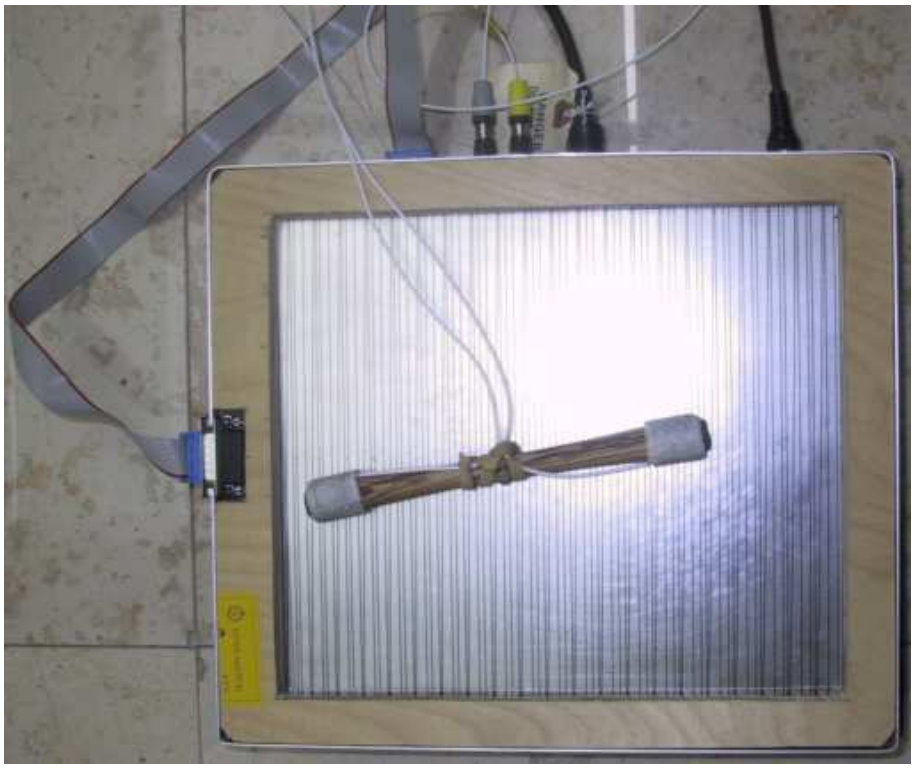


Figure 3.1: The *Radiobaton* controller with sensors connected to a Bodhran stick.

consists of a base–frame (compare also figure 3.1), in the case of the *Invisiball* placed below the elastic cloth, and small transmitters, whose position in three dimensions is tracked (figure 3.3).

Figure 3.2: The Invisiball; the elastic cloth that can be pushed down with the sensor connected to a finger(see also figure 3.3) represents a surface where a virtual ball is rolling on.

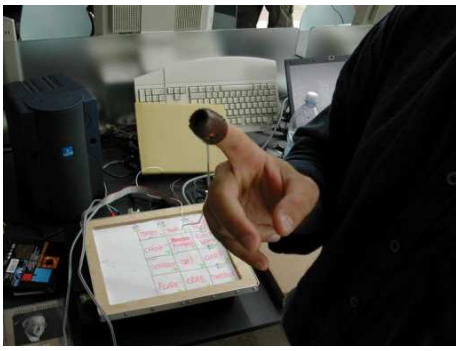


Figure 3.3: The finger sensor of the *Invisiball*.

The final practical implementation of the *Invisiball* showed to be problematic, as little inexactnesses in the measurement (of the control gesture) and the computation of the resulting movement of the ball added up to a somewhat awkward feeling of the device as a whole. Further on, the directly perceived, seen and felt, deformation of the representing surface was hard to synchronize (a.o. due to the restricted resolution of the sensors) with the virtual surface shape and movement, resulting in a certain mismatch between different perceptual channels. In the context of this thesis the *Invisiball* was important since the collected experiences revealed important points that were taken into account in the development of the next tangible–audio-visual interface, the *Ballancer* presented in the next section (3.3). It was seen that even small and occasional "misbehavior" of the device, with respect to the expectations raised to a user on the background of everyday experience, can strongly derogate convincingness and a user's contentness. For the *Ballancer*, strengthened attention was thus paid from the beginning, that the whole chain from the acquisition and measurement of the gestural control movement to the computation of the resulting virtual behavior could be practically realized with high exactness. To achieve a sufficiently strong match between the experiences through the different perceptual channels and the direct feedback from the control access (versus the feedback about the virtual scenario), a much more simple, "robust" (as compared to the *Invisiball*) control metaphor was chosen for the *Ballancer*. Appearing as rather unspectacular in itself, the simple principle behind the *Ballancer* (balancing a ball on a tiltable track, section 3.3) showed to be very strong in its clear, stable practical realization, in terms of usability, expressiveness and assessment (chapter 4).

3.3 The *Ballancer* metaphor and interface

The last interactive multi-modal system constructed during the course of this work is again integrating the rolling-model, as the most complex, versatile and expressive of the developed sound models, into a larger metaphor of function and control, together with a tangible input device and visual and sonic feedback (namely, basically of a rolling-sound). With the experience from the *Invisiball* in mind (section 3.2), here a particularly simple overall metaphor was chosen, that of balancing a ball on a tiltable track. The (virtual) ball is free to move along one axis over the length of the track, being stopped or bouncing back when reaching the extremities. The acceleration of the ball along the length of the track is directly related to the vertical angle. More exactly, if the track forms an angle α with the horizontal plane, the acceleration a along the track results from the vertical gravity acceleration g via

$$a = g \cdot \sin(\alpha) \ . \tag{3.1}$$

Any effects of the changing vertical ball position induced by tilting the track are neglected. Further, all damping of the ball movement through friction on the track and in the air is modeled by one term of friction force f, proportional

to instantaneous velocity v (in the direction of the track length):

$$f = -k \cdot v \ . \tag{3.2}$$

Finally, in considering the ball displacement along the track, all effects of rotation, such as the moment of inertia, are ignored. The position x of the ball on the track is described by the following differential equation:

$$\ddot{x} = \sin(\alpha) \cdot g - k \cdot \dot{x} \ . \tag{3.3}$$

The rationale of the system metaphor is substantiated by the following points:

- The simplicity of the idea supports a robust realization. The lesson from the *Invisiball* (section 3.2) that is painfully sensitive to practical imperfection (ranging from the exact definition of the movement of surface and ball to the detection of the controlling finger) has been learned.
- The general principle of the balancing-metaphor, as well as its haptic control, is familiar from everyday experience. It is thus easy to understand for an average user, even without explanation and after very little training [2].
- The control movement of the user is, in its repercussion on the system behavior (via the movement of the virtual track and ball), concentrated in only one (one-dimensional) variable, the track angle. This is a great advantage for in–depth evaluation, as reported in the next chapter(4); in fact the most far-reaching results of the evaluation would probably not have been possible without such a clear and precise representation of the control movements.
- Working on the same general balancing notion, the system could easily be expanded, e.g. to a two-dimensional plane.

[2]The earnest, objective authorization of this statement is one of the results of the following evaluation (chapter 4).

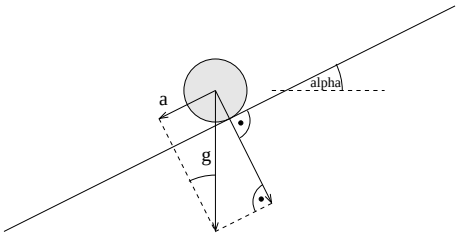


Figure 3.4: *Scheme of a ball rolling on a tilted track. The gravity acceleration is split into two terms parallel and perpendicular to the track, according to the track-angle.*

- The metaphor can be adapted to a wide range of control tasks. The system can be seen as a possibly simple representation of a controlled system that is reacting with non-negligible inertia. The notion, that is important for the wider interpretation of the setting of the *Ballancer*, also in its relation to classic settings of Fitts' law, is explained together with an according test task (see 4.3.1) in the next (evaluation) chapter.

Another strong advantage is that the physical, purely mechanical realization of the metaphor is straightforward. For instance, in the practical implementation the control track can also hold a real ball moving on its top surface. In this way the virtual system can be directly compared to its mechanical pendant, to measure how far it is from the "real thing".

3.3.1 Implementation

The complete software part of the tangible-audible interface is again realized in *pd* [55] and runs with low computational load [3] on a standard personal computer.

The ball equation of motion (3.3) is transferred into discrete time, at a rate (in the range of 100 Hz) much lower than audio sampling rate. The resulting calculation as well as higher-level structures of the rolling-model are defined by means of the *pd*-GUI.

A schematic graphical representation of the balancing track and the rolling ball is implemented in *gem*[4], an *openGL* rendering-extension for *pd*.

The interface is physically controlled by holding and tilting the rolling track, a 1.05m wooden bar. This tangible controller has an aluminium track attached

Figure 3.5: The "rolling-track" with a glass marble rolling in its upper-face aluminium track.

[3] A certain exception is the graphical representation mentioned below.
[4] http://gem.iem.at

to its upper face, which can hold (e.g.) a glass marble of 2.5 cm diameter rolling along the track according to the tilting-angle.

Fixed to the rolling track is an accelerometer that measures acceleration in the direction of the length of the track. This measured acceleration is the fraction of gravity in this direction, as described in equation (3.1). We can thus calculate the tilt angle from the accelerometer output, again using the *pd* environment. The data-transfer from the (analog) accelerometer to the software is established through a *Kroonde*[5] sensor wireless interface, connected to the computer via a UDP socket-connection.

[5] http://www.la-kitchen.fr

Chapter 4

Evaluation of rolling model and *Ballancer*

4.1 Introduction — General Considerations

The general point of the user evaluation tests described in the following is the demonstration of the informative potential of the sound model of rolling for human–computer interaction, on the basis of *ecological perception*, as motivated and displayed in chapters 1 and 2. The conveyance of information through the synthesized sonic feedback and its exploitation by users, in the sense of ecological perception, its therefore immediate intuitive application, is proved and illuminated. At the same time, the results of the experiments may form a valuable contribution to the knowledge about the mechanisms of perception and exploitation of continuous acoustic feedback by humans in everyday scenarios.

The validation of the sound model reported in the following is partially unavoidably interconnected with a parallel evaluation of the implemented control interface that is used during the largest part of the test. This interface, the *Ballancer*, however is only one possible choice for the control of the rolling sound model which is the ultimate point of focus here — or more general, the exploitation of continuous acoustic feedback by humans in situations of *everyday-listening*, for which the rolling model forms one carefully chosen representative.

For the purpose of clarity I would like to structure the possible processes of acoustic conveyance of information [1], as considered relevant in this context, into three general categories.

- The first category, referred to in the following as "*sound identification*", relates to the notion that sound events(or also -models) can be informative [2] by virtue of the capability to provoke the (sufficiently clear, reliable) connotation with a known general familiar scenario. Concretely, in this

[1] The term "information" is here used in a possibly wide sense.
[2] . . . in a sense sketched in the following sentences,

category fall the questions if, and how coercively, the sound of the model validated here is identified as "rolling" and how the perceived objects and actions are characterized, e.g., what structure, size or material of the surface and rolling object are perceived. If these questions can be answered positively with sufficient unambiguousness the sound model can inform a user in an intuitive way, i.e. without dedicated training or conscious explanation, about the class or nature of an interactive system and serve to determine and steer his way to approach and interact with the system.

- As noted in chapter 1, sound can continuously and instantaneously reflect properties of ongoing processes (*transformational attributes*), e.g. of user-interaction. The sound of rolling generally reflects the speed of the rolling object and possibly its direction and position. The important point here is to show that users do perceive and understand this information and further on make use of it (possibly without awareness), reflected in "*performance improvement*". [3]

- Sonic feedback may also convey information in a wider, subjective, "*subject-centered*, emotion-related sense that is harder to capture, formalize or quantify. It can create or enable the feeling of *presence* in a virtual/augmented environment (see e.g. [21], [34]), provoke engagement and raise interest of a user and increase his feeling of comfort and confidence while interacting with a system.

Of course these three areas are to be seen as (hopefully) supportive rather than absolute constructions, and overlap and mutually depend. E.g. may it be expected that a registered improvement of performance that is reported to a user will improve his subjective feeling of comfort with and confidence in the system/interface he is controlling/interacting with, and vice versa. Similarly, raised engagement or motivation to approach and deal with a system or interface might have a positive influence on achieved objective, measurable performance values, and a better recognition of an employed control metaphor through an appropriate sound component may in turn improve a user's motivation and subjective contentment.

The following evaluation experiments mostly focus on the second field of *performance improvement* through sonic feedback, that is considered the central "hard fact" which can be formulated and measured in the clearest, most doubtless way. Also, this point and the underlying mechanisms seem to have been examined to the least extend before and are of direct, unquestionably high relevance for the possible use of sound models in human–computer interfaces. The point of *sound identification* is addressed in a shorter initial part of the evaluation test consisting of listening- and controlling-trials and a questionnaire. Not directly dealt with is the last complex of *subject-centered* evaluation, not least because this aspect is hard to formalize. User feedback during the performance tasks and especially additional verbal user responses however allow

[3] One concrete meaning of "improved performance" is cleared in the respective paragraph explaining the experiments and results.

to draw some first preliminary conclusions and point out directions for possible dedicated examination of this aspect.

A total of ten subjects participated in the evaluation. They were six men and four women of age between 23 and 32, chosen among students of electrical engineering at the Technical University of Stockholm (KTH, Kungl. Tekniska Hoegskolan). Subjects had no previous awareness or knowledge whatsoever of the work presented here. Each session (one subject) consisted of a shorter (ca. 20 mins), listening/trying- and interview part, dealing with *sound recognition* and a longer (ca. 1h) performance part addressing the aspect of *performance improvement*. Subjects were told to feel free and give any remarks coming to their mind during the tests, also without being asked explicitely. Some of these free spontaneous comments are interesting and revealing; they are cited at appropriate positions in the following description and discussion of test results. For their participation in the experiment, subjects were paid 80 Swedish Crowns (ca. 9 Euro) each.

4.2 Sound recognition and understanding of the metaphor

To examine which (if at all) spontaneous connotation the sound generated by the rolling model provokes by itself, subjects were at the beginning of their testing session played two short sound examples of the model, each followed by the question "What do you hear?". The sounds were presented through headphones without previous information about their origin and background whatsoever. Both sounds were generated with parameters according to a small, hard ball of 2.5cm diameter rolling in right–to–left direction on a hard surface rather fast in the beginning, then subsequently slowing down to a stop as if being rolled on a horizontal surface. One of the two sounds also contained a few accelerating initial bounces as if the ball was being dropped on the surface and rebouncing for some period before finally rolling. This initial bouncing phase was generated with the bouncing model described in chapter 2 with resonator parameters set identically to those in the rolling model. To provide the possibility of eventual repetition of the identification task, these test sounds can be downloaded from the author's webpage [66]. The motivation of the choice of two sound examples was to test if such a typical starting dropping incident contributes to the identification of the scenario. Previous informal experience had suggested this conjecture. The order of presentation of the two sounds (with the subsequent question) was varied, "no-bouncing – bouncing" for one half of the subjects, opposite for the other five. Table 4.1 gives an overview of the answers. [4]

Next, blindfolded subjects were given access to the balancing-track and asked to carefully move up and down their arm holding the track. Testing the device

[4] I cite subjects' use of the term "same", referring to the answer to the immediately preceding question/setting.

in its sonic reaction to their movement for as long as they wished, subjects were asked to identify "What is going on here?"; table 4.2 shows the answers.

For both, the test sounds as well as for the sonic feedback of the *Ballancer* the direct output of the model was taken, i.e. the modeled vibration of the

subject	associativity of rolling sound	
	synthetic without initial bouncing	synthetic with initial bouncing
1	"small ball going from right to left"	"same ball, dropped, then rolling and jumping a bit"
6	"small metal ball rolling from right to left across some hard surface"	"small metal ball rolling, this time more egg-shaped"
7	"small, hard, like iron, ball, diameter ca. 2cm, rolling on a smooth and hard surface; some small dips right from the middle"	"similar as before, dropped in the beginning"
8	"hard ball, steel or glass, diameter ca. 3cm, rolling on a hard, e.g. marble surface"	"like before, a little bit smaller, because 'wiggling' more"
9	"steel ball rolling on a hard surface, diameter $1 - 1.5$cm"	"about the same as before, bounces in the beginning"
2	"rolling ball going in circles, fast in the beginning then slower, like rolling up a drain, from right to left"	very first impression: "squeaking door", then correction: "ball that falls down and then rolls; hard, e.g. marble ball on marble surface; size about 3cm diameter"
3	"metal ball rolling in a bowl"	"starting engine, impulses of increasing frequency merging into a continuous sound"
4	"kind of metal ball rolling"	"ball falling, bouncing and then rolling away"
5	"rolling object, this time not falling"	"some kind of ball (like a ping pong-) bouncing and then rolling"
10	"ball rolling from right to left, about the same size"	"ball bouncing and then rolling from right to left, diameter ca. 3cm"

Table 4.1: "What do you hear?" — answers of the ten subjects after listening to a synthesized rolling sound, without and with (presented in this order for subjects 1 and 6 to 9, therefore the swapped ordering in the list) initial bouncing

object at one point without any consideration of spatial sound propagation. [5] As the only form of post-processing, the right–left movement was acoustically displayed through simple amplitude panning.

In a second step, the previous test procedure was repeated, this time with a real glass marble of ca. 3cm diameter rolling on the track (replacing the virtual ball and synthesized sound). Blindfolded subjects were made listen to the sound of the small marble and again asked "What do you hear?"; they were finally given access to the track as before followed by the same question of "What is going on here?". Answers are shown in table 4.3.

4.2.1 Results

Questions and respective answers in the first, *sound/metaphor identification*-, part of the evaluation are shown in tables 4.1 to 4.3. The main results are:

- Overall association of the synthetic sound with rolling was very high: All 10 subjects identified the sound example without initial bouncing as a

[5] A mechanical pendant would be the signal as picked up by a contact microphone.

sub-ject	associativity of audible–tangible device, virtual realization
1	"small ball rolling in a tiltable pipe, bumps at the end"
2	"ball rolling in a tiltable pipe according to the angle; different surface texture somewhere near the middle, some kind of bumps; rebounces and eventually stops at the ends"
3	"sound of a wave plus of a metal ball rolling on a track that I'm tilting"
4	"I'm holding a tube with a ball inside that rolls towards the end that is lowered; smooth surface, but irregularities near the middle"
5	"ball rolling up and down a pipe that I'm tilting; obstacles near the middle"
6	"same ball rolling in a ramp that I'm tilting, holding at one end"
7	"small ball as before, I'm controlling the angle of the surface"
8	"I'm controlling the tilt of a surface where the metal ball is rolling on; near the middle rougher surface (like asphalt versus marble), bumps or stripes"
9	"I'm holding a tube that's fixed somewhere, with a ball rolling inside from side to side; slightly right from the middle a rougher area"
10	"ball rolling on a plane or in a tube that I'm tilting; near the middle section with bumps"

Table 4.2: "What is going on here?" — Blindfolded subjects' answers when accessing the *Ballancer*

rolling ball. Surprisingly, after the informal expectations mentioned above, the identification of the sound example with initial bouncing was slightly less clear. One subject described the sound as coming from a starting engine, another subject spontaneously mentioned a squeaking door before changing its mind (without any hint or additional question by the experimenter) and stating a falling, then rolling ball. Also interesting, these two outliers belong both to the group that was presented the "bouncing plus rolling"-sound first. It may seem that the decidedness of the connotation provoked by this sound gets stronger when subjects had already heard the other, rolling–only, sound. In fact, previous informal experiences had suggested the exact opposite effect: it was expected that the bouncing

subject	associativity of rolling sound — mechanical event
1	"smaller balls or cylinders, a couple of mms in diameter (or maybe bumps make one sound like several)"
2	"same scenario as before, maybe more than one ball; smaller, little less than 1cm"
3	"smaller balls, maybe 2, connected, rolling in a track"
4	"couple of small balls rolling on a pipe, diameter ca. 5mm"
5	"something like a toy car being moved/pushed"
6	"something rolling, diameter maybe 5mm"
7	"small ball, diameter ca. 8mm, rolling in a pipe"
8	"hard object sliding in a grove, or a ball rolling inside a tube/pipe"
9	"some sort of ball in a tube, smaller, diameter ca. $2-3$ mm"
10	"wheel going from side to side in a track"

subject	associativity of audible–tangible device — mechanical realization
1	"bigger objects, or maybe one bouncing several times"
2	"as before, maybe more than one ball"
3	"small ball, or maybe two, rolling in a track that I'm holding at the end"
4	"like before"
5	"something like a marble rolling up and down a surface/pipe"
6	"several (two or more) objects rolling, connected to each other"
7	"as before"
8	"two metal balls inside a tube, diameter ca. 1cm"
9	"as before, seems larger at the ends, maybe 5mm"
10	"wheel or ball in a track that I'm tilting"

Table 4.3: Identification of the mechanical scenario by the 10 subjects, from the sound only (above, "What do you hear?") and when accessing the device, blindfolded (below, "What is going on here?")

event, which is clearly of much more simple nature, would help to create the impression of a ball and thus support the recognition of the rolling event. Finally, two subjects described the difference of the +bouncing example as compared to the rolling–only sound in terms of the shape ("more egg-shaped") resp. the size ("this time smaller, because 'wiggling' more") of the rolling object. A distinct evaluation of the identification of this bouncing event would surely be a useful addition in this respect.

- The sound of the small glass marble rolling on the track in front of blindfolded subjects turned out to be more ambiguous than the synthesized sounds (at least the rolling–only example). Only 3 subjects clearly stated one rolling object, while 3 subjects heard *several* objects rolling simultaneously and one other was not sure about the presence of one or several rolling objects. One subject was not sure if the object was rolling or sliding, one subject heard "a wheel" another "something like a toy car". Yet another test subject heard the ball *inside a tube* and another mentioned this possibility.

- When controlling (blindfolded) the tangible–audible device with the synthesized sound feedback, all 10 subjects clearly described an object rolling on a surface whose angle is controlled by tilting it around some fixed axis. Only one subject mentioned an additional "wave"; the same subject (no. 3) is also the only real exception in the recognition task with the synthesized rolling sound. Two subjects described the object rolling inside a tube which appears to be a rather cognitive decision based on the fact the object does not fall out of the track, since the same persons did not (in fact no subject did) make this description when only listening to the sound of the same model.

- The ambiguity in the (purely auditory, blindfolded) perception of the mechanical scenario did not diminish when subjects were given access to the track and were allowed to control it. Remarkably, the identification of the scenario changed for some subjects when they were allowed to control it, but overall the recognition of the de–facto scenario did not improve. Concretely, only one subject (no. 5) got closer to the de–facto setting in its description when allowed to control the track herself, while another (no. 8) was further misled in that case.

Some additional remarks have to be made concerning the results of the recognition tests:

- Also the diameter of the (real) glass marble was regularly guessed much smaller than its de–facto size of 3cm, between 2mm and 1cm. The size of the virtual ball instead was described to lie between 1-1.5 and 3cm, much closer to the intended diameter of 2.5cm. It has to be remarked that some subjects made this guess spontaneously, which raised the idea to the experimenter to explicitely ask following subjects for a guess. This request was not made to all subjects and it was not protocoled if regarding values

where given spontaneously or on demand. In that way, the handling of subjects was in that point not perfectly planned and consistent which is surely a potential point of criticism that might be stabilized in a future test. I nevertheless decided to include these informal results concerning size into this discussion as interesting but without weighting them too high as proofs for essential claims.

- The same remark as the previous has to be made about subjects' statements concerning details of the surface, in their reactions when accessing the virtual tangible–audible device.
- The sound model of rolling is the result of a process of abstraction and *cartoonification* that has extensively been displayed in previous chapters(1, 2). The model is by no means meant to be a possibly perfect simulation of one specific individual mechanical scenario. In particular does it not specifically correspond to the mechanical version of the *Ballancer* (with the glass marble) which is just one possible mechanical realization of the general metaphor. Of course a functionally equal mechanical device realized with a different combination of materials in a different construction might be less ambiguous and misleading in its acoustic appearance.

Summarizing the results of the questions about the sounds and the tangible–audible device, it can be said that the modeled sound and metaphor are intuitively understood. In this way the sound model has an informative meaning by itself without additional information or explanation and obviously is very usable to accompany and support representations of rolling actions in other perceptual modes, e.g. visual or tactile. The combination of modeling everyday sounds and using a familiar control metaphor here exhibits the advantage that virtually no explanation and learning are necessary. As opposed to what happens with abstract sounds/controls [15], users may immediately understand and react to transported information without being instructed. The spontaneous impression of the intended scenario (rolling) is even more clear for the tangible–audible interface than for the compared mechanical device that provides a physical realization of the metaphor. This demonstrates how effective the *cartoonification* approach to sound modeling can be: although the device is perceived as fictitious [6], nevertheless it can very reliably elicit an intended mental association, even more clearly than certain realizations of the "real thing". [7]

4.3 Performance measurement

While the first part of the evaluation test has shown that users understand, i.e. identify and accept as convincing, the sound model of rolling and the *Ballancer*-interface, the second part addresses the question if users also appropriately use

[6] . . . not least due to the absence of any spatial sound propagation,

[7] Here the previous remark has to be kept in mind: I do not claim that the sound model is more reliable in its provoked connotation than *any* mechanical realization of the metaphor but to "outperform" some real objects (at least the chosen example).

the device and perceive dynamic ecological attributes contained in the sound and exploit this information. These dynamic attributes are generally the position, velocity and direction of movement of the ball and, related, the local structure of the rolling-surface at the momentary position of the ball (to be exact, in the direction of its movement). Size, weight, hardness and *sphericity* [8] of the ball could also change dynamically but are fixed in the following performance test; in fact, also in physical reality the situation of a rolling object that changes in form and size during the movement is rather unfamiliar. As a result of the choice of a possibly simple control metaphor, with the ball moving only along one dimension, the direction of its momentary movement is restricted to "right" or "left" and has no influence on the tracked surface profile in this case. The following part of the evaluation thus isolates the most important dynamic attributes, the position and velocity of the ball. During the movement of the ball, the surface profile at its momentary position is constantly reflected in the sound; vice versa, the emitted sound informs about the ball's position on the surface, since the latter does not change in form. More exactly, after the above description of the *Ballancer* it can be heard [9] if the ball is currently moving inside or outside the target area. In particular, an abrupt change of the surface structure, further underlined by a little step due to the different depths of surface profiles marks the moments when the ball enters and leaves the target area. As the model does not consider any spatial sound propagation, the momentary position of the ball is further on expressed only through simple stereo amplitude panning between left and right. From this behavior of the sonic feedback, the position of the ball can be perceived with much less precision than it can be perceived in a good [10] visual representation. The following tests show that subjects however do generally understand the position information contained in the sound of the *Ballancer*, at least to the extend necessary to perform the test task with purely auditory feedback.

Velocity is here considered the main attribute of interest: I conjecture that this parameter can be *heard*, perceived acoustically, more "direct" [11] than visually. The point to note is here that velocity can generally visually only be extrapolated from the perceived position over time. To clear this notion, one might imagine a momentary visual glimpse, e.g. a photo of a moving ball: it is not possible to judge the speed of the ball at this moment. Further information is necessary to depict the velocity of an object, e.g. in form of a blurred picture which is one form of integrating information during a time span, or through arrows in a graphical representation. On the other hand, momentary velocity *is* constantly reflected in the sound of a rolling object. More exactly, I would here have to define precisely what is a "momentary sound". Of course I am again

[8] The meaning of this parameter is explained as part of the description of the rolling model in chapter 2.

[9] This informal experience is proven by the results of the following tests.

[10] E.g. in comparison to a graphic display spanning a standard computer screen... It is not the subject of this text to further specify the quality of graphical displays or quantitatively compare the resolution of position in possible graphical displays with the sound model.

[11] ... in a sense explained in the following,

talking about a short period of an acoustic signal — sound only exists in time —, but the claim addressed and supported by the following test is that "momentary" acoustic perception of velocity is faster, more precise [12] than its visual perception. In other words, it is postulated that momentary velocity is acoustically perceivable with higher temporal resolution — from very short segments of a signal — and that human listeners, or operators, exploit this continuous information.

Perhaps the most obvious way to examine the perception of rolling-velocity from the sound would be to directly ask subjects about the velocity while listening to different generated sounds. One might think of a sorting- or scaling task with generated sounds of various velocity or explicit questions about the (development of the) velocity of an acoustically modeled rolling object [35]. There is here the problem that answers might reflect a conscious reaction of the test subject on the question rather than a spontaneous perception. E.g., it is possible that subjects connect a sound with a faster modulating amplitude to a faster moving object when they are suggested (or even forced) to make a choice, although they would not spontaneously have this connotation without being asked. In fact I believe that such processes of perception may often be out of subjective awareness and thus hard to verify through questions resulting in a conscious answer. The approach here is therefore somewhat indirect and more complicated. Subjects are asked to perform a specific control task and their movement while solving the task with and without acoustic/graphic feedback is recorded and analyzed. From systematic differences in the subjects' movements under the different sensory conditions it can be concluded that the information they (the subjects) perceive depends on the stimulated sensory channels. Through deeper analysis of the control behavior with and without sonic feedback I can finally support the conjectures given above. This indirect strategy of using a performance task allows to illuminate processes of perception and human information processing that the subject may not be aware of, without biasing through cognitive questions/stimulation. In fact, reactions during the tests underline the unaware nature of the process. Further on, besides proving the absorption of different sensory information, their immediate active exploitation in human control gestures can be demonstrated, and some useful [13] quantitative measures be given.

4.3.1 The task

In order to examine if and how subjects perceive and use information through different sensory channels about the movement (position and velocity) of the virtual ball on the *Ballancer*, they were asked to perform a specific control task with various configurations of sensory feedback. The task consisted of moving the ball from a resting position at the left end of the balancing-track, held horizontal at the start, to the target area of 15cm length slightly right

[12] . . . at least from sounds as e.g. rolling under the preconditions of the experiment,
[13] . . . e.g. in concrete implementations,

of the middle of the track, and stopping it inside here. On the mechanical representation, the target area is marked with black adhesive tape (as seen in the photos of figures 3.2 and 4.2). The boundaries of the target area were located 10cm and 25cm right from the center, i.e. 60cm and 75cm from the left end of the track. Subjects were asked to try and accomplish the task as fast as they could and the needed "task time" was measured. More exactly, the task was counted as fulfilled when the ball stayed at rest (no movement) inside the target area for at least 1 second; the time was measured from an acoustic starting signal (a "ping" sound, given with the ball resting in initial position) to the first successful stop as defined. [14] The seemingly complicated criterion of task completion is necessary to guaranty unambiguous measurements, since under the conditions of control here, the ball will stay at rest only with a minimum of attention/concentration from the controlling subject.

Relation to settings of Fitts' law and wider context

The performance task in the second part of the evaluation test, moving the ball from a given distance into a fixed target area and stopping it therein, is analog to the classical settings of Fitts' law tests [24]. In Fitts' original experiment, one of the most important, central works for human–computer interaction and the modeling of human movement, subjects had to move a stylus over given distances into a target range. In conventional Fitts–like experiments however (see e.g. [47]), the moved stylus (if used) is of basically negligible weight and size compared to the (mass of the) human arm and can be seen as a marker connected to the human body: Fitts' law applies to human movement as such, generally not directly to interaction. Human movement in itself often causes rather low acoustic feedback, the latter is mostly a consequence of the resulting interaction with and of external objects. Accordingly, the question of sensory feedback in such movement tasks has not been followed deeper so far; in classic settings of Fitts' law subjects simply see and feel (proprioceptive feedback) their own movement.

Many tasks of interaction, with computers as well as in everyday surroundings, however are quite different from the isolated human movements as examined by Fitts, due to significantly different conditions of sensory, and force-, feedback. Objects or systems that are controlled by a human *operator* [15] often react with non-negligible inertia, in such a way that the *operator*'s input movement is more closely related to the resulting acceleration (of the object or system), rather than purely its position. Similarly, the size or range of a controlled object may strongly exceed the dimensions of the *operator*'s body. As a consequence, the question of how the *operator* receives feedback about the movement of the object or system he is controlling, gets highly relevant. When a heavy object is grabbed and moved directly, e.g., its (visual) position may be

[14] It is these measured times and the statistical distribution of these measurements, for individual, or groups of, subjects, what is comprised by the term "performance" that has been used before without clear definition, trusting in a rather intuitive understandability.

[15] I here use this term that is common in robotics.

parallel to that of the *operator*'s body and the applied acceleration may be felt immediately by the *operator* (force feedback). But often, control interaction is much less simple and even these conditions of direct visual and tactile feedback are not given. A familiar everyday example that can serve to illustrate all the preceding considerations is that of driving a car, where the position of the driver's foot on the gas pedal (or brake) relates to the acceleration of the car and not directly (e.g. proportionally) to its position; further on, the force feedback from the pedal as felt by the driver is clearly not the force acting on the car, generally not even proportional, and as the driver himself is located inside the car he visually controls its position not relative to his own standpoint but relative to the surrounding environment. On the other hand, acoustic feedback of car movement, originating from the motor, wheels..., appears to convey valuable information for the driver, but such possible effects of continuous auditory feedback have never been formally verified or assessed.

Based on the considerations above, the task in this test isolates some of the main aspects described. In balancing the ball on the track, the vertical position of the controlling hand(s) (more exactly the vertical distance of both hands) is approximately proportional to the acceleration of the ball. This is a consequence of equations (3.1) resp. (3.3) (section 3.3) and the consideration that the damping of the movement is comparatively small for larger control movements. The balancing task here, thus isolates "acceleration control" as sketched above in the "driving" example but in the most simple context of one-dimensional movement (along the track). In accordance with the mentioned common unavailability or "unreliability" of force feedback in complex control tasks, the *Ballancer* does not give force feedback whatsoever about the movement of the controlled ball. Basically, the *Ballancer* exerts very small resistance forces at all, as the inertia (mass...) of the control track is small compared to the controlling human arms. This somewhat idealized situation has been chosen here, since it gives particular relevance to the question of auditory and visual feedback. The latter is very simple here, a schematic display of adjustable size representing the virtual track and ball, and auditory feedback is of course at the center of interest. I like to note that the described focused nature of the *Ballancer* setting and test task is achieved without getting "clinical": the previous and following results of evaluation underline the familiarity and "natural" understanding of the device and task.

Besides the specific central points given so far, the setting in the second part of the test is also relevant and adaptable for concrete applications. Common tasks of steering, navigation or control may be formulated in terms of reaching and holding an equilibrium. The balancing metaphor may be useful wherever direct position control is not suitable, such as in portable devices where navigation by tilting has been suggested and used [23][25].

Finally, it has to be noted that in the following experiment the dimensions of the target area and the "target distance" are fixed, in contrast to Fitts' experiment. The first goal here is to detect and illuminate as such, effects of auditory feedback. The attempt to derive a model or rules for control movement analog to Fitts' law would be a possible next step. The following experiment can be seen

as a try to establish basic knowledge about interaction with auditory feedback, parallel to the familiar, intuitive proprioceptive basis that Fitts' examinations start from.

Experimental design

In individual sessions, the ten subjects, after having absolved the test of *sound/-metaphor recognition* reported in the previous section 4.2, were asked to perform the task described above under different conditions of sensory feedback described below, and told to try and be as fast as possible. Subjects were not informed anyhow about their measured times needed in the trials, in order to minimize effects of conscious adaptation to the test conditions and isolate the effects of mechanisms applied by the subjects without awareness, trying to optimize (subjectively) their performance. Movements of control and ball, i.e. the changing angle of the rolling-track and the position of the (virtual) ball during trials, were recorded for later analysis. Figure 4.1 shows typical recorded trajectories of the ball during task performance. [16]

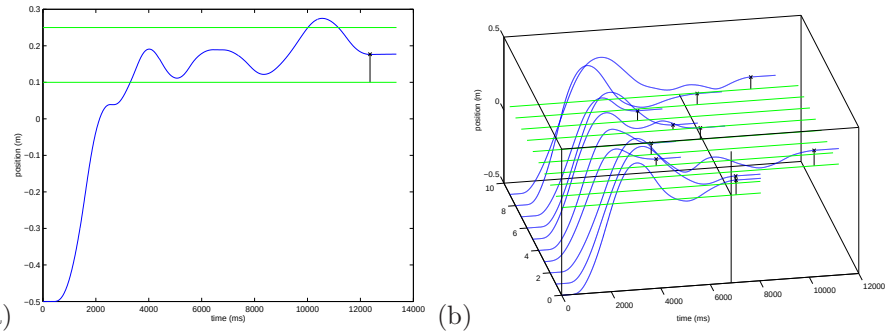


Figure 4.1: Recorded trajectories of the virtual ball during performance of the task, a single trial (display size 2, no sound) and a whole set of 10 trials (subject 8, largest display, no sound). The ball starts from the left end of the track, 0.5m left from the center, enters the target area 10cm right from the center (indicated by the lower green line) and finally comes to rest inside the target area for at least 1s, marked by the black "⋆" and vertical line (task completion — The horizontal and vertical black lines in figure (b) mark the average task time.). In less efficient trials, the ball may temporarily leave again the target area on the right side (upper green line in (a)).

Feedback about the movement of the virtual ball during the trials was given acoustically through sound from the rolling model (chapter 2) played over stereo headphones, and/or visually on the computer screen, as a schematic representation of the ball on the track (see figure 4.2). The graphical display, with the ball represented as a monochrome (red) sphere on a line representing the track

[16] The display size factors are explained below.

and the target area marked by a different color (light green), was realized in 4 different sizes. Scaling factors for graphical display were ranging from 12, the largest, with the track horizontally filling the 21" computer screen (as in figure 4.2), over 4 and 2 to the smallest, 1. In the latter, smallest, size, the moving sphere (representing the ball) could not always be visually detected due to the boundaries of the screen resolution [17], 1024×768.

Each test started with 2×10 training runs (10 plus "pause" plus 10) with the largest display ("full screen", scaling factor 12) and sound feedback, to minimize possible training effects. Subjects were told that these first 20 trials could be used to get familiar with the setting and practice the task. In the following runs, the needed time was measured with display sizes of 12, 4, 2 and 1 (in this fixed order); again 20 measurements were made for each size, 10 times with and 10 times without sonic feedback. The order of the measurements "without–with

[17] Details of this last condition of the smallest display are not further described here, since the according results are finally not considered, as described exactly below.

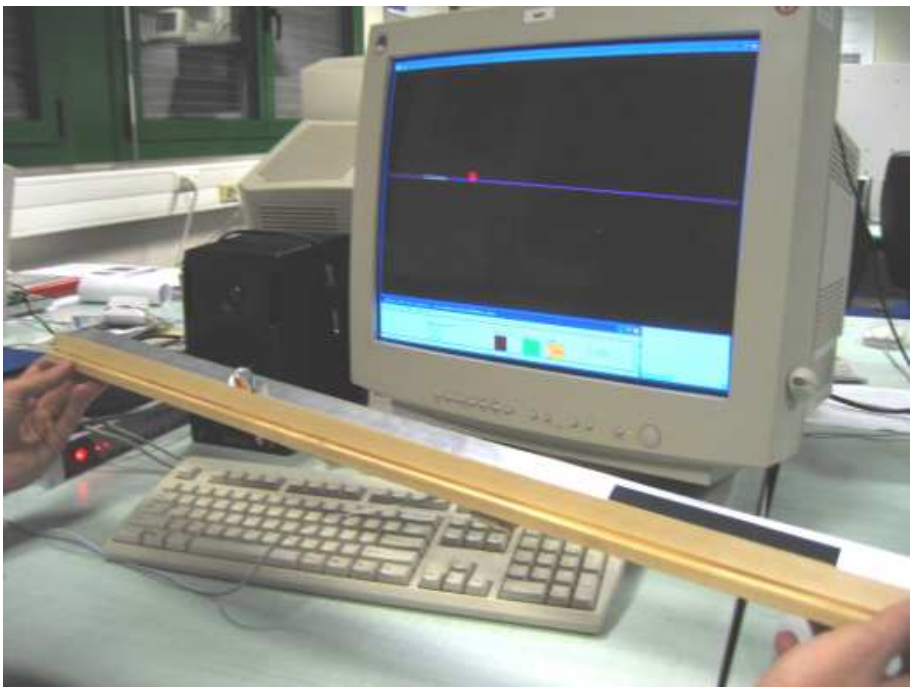


Figure 4.2: The *Ballancer* with the graphical spanning the whole 21" monitor (display factor 12). The photo was taken during use of the device in a game application, thus the green target area on the screen is not in the same fixed position as during the performance task described here (slightly right from the middle, according the black mark on the mechanical track). Also, the real glass marble on the track only serves demonstration purposes in the photo.

sound" resp. "with–without" was switched after half of the subjects to test for, and eventually counterbalance an effect of the order of performance on the results. At each change of the display size subjects had a short rest [18] and were afterwards given an additional 3 trials to warm up under the new conditions before the start of the actual measurement. Finally, the display was fully closed and subjects were asked to try and perform the task only with sonic feedback.

It has to be noted that the work presented here is concerned only with auditory display and partly its possible interaction with other sensory modes. The different visual conditions, i.e. different display sizes, are to be seen as different, *independent* background settings for the examination of effects of the auditory feedback. It is *not* the goal of this work to mutually compare the different visual settings, and the results of the following tests should not be seen and used as measures of the effects of varying display size. The order of presentation of the various graphical displays would have to be taken into account and possibly randomized or counterbalanced to really receive solid insights concerning the size of graphical displays. [19]

4.3.2 Results

Informal preliminary experimentation (of the author and others) with the *Ballancer* had revealed that it is possible to solve the "target-matching" task described above with purely auditory feedback. [20] On the other hand it was found that with a sufficiently big graphical display, e.g. scaling factors 12 (whole screen), 4 and 2, the task is solvable without sound, much more easily [21] than in the sound–only configuration. Again from subjective experience in the informal tests, for the larger display sizes, scaling factor above ca. 4, additional auditory feedback did not seem to clearly alleviate the difficulty of the task, as compared to purely visual feedback [22] . Finally with display sizes smaller than ca. 1 the schematic representation of the rolling ball is not always clearly perceivable, depending on momentary angle and position, due to the boundaries of the screen resolution; it showed to be very difficult to solve the task only with visual feedback from such a small display, in fact only with a certain amount of guessing and trying, partly more like a game of luck. [23] From

[18] A short pause was needed by the experimenter to adjust the new display (and other connected) settings.

[19] Of course I presume that a reduction of the display size means a decrease of the available visual information, which should influence the significance of additional or alternative auditory display. But the tests are not designed to quantify and further substantiate this notion. It was further on not a goal to directly compare visual and sonic feedback. The exact reasoning behind the developed test setting with different display sizes is discussed together with expected and finally examined results in the following section 4.3.2.

[20] In fact the dimensions and position of the target area have been chosen "near the boundary of solvability" for the sound–only task.

[21] from a subjective standpoint...

[22] Some test subjects later stated the same subjective experience, as I will describe at the respective point of the detailed discussion of the test results below.

[23] In the course of the tests, trials with the smallest display turned out to be problematic because of artefacts of the low screen resolution, but not essential for the main results on the

these preparatory observations it could be suspected that at least under certain conditions of visual feedback, here certain rather small display sizes, additional auditory feedback could support the solvability of the task and improve the time needed in subjects' performances. As already noted, it is not a goal of the tests to compare the different display sizes and clarify the obviously expected performance improvement with increasing display size. Nor is it a direct goal to compare performances with purely visual or purely auditory feedback. To that end it would be necessary to counterbalance the order of presentation of the various sizes, which would in turn strongly suggest more subjects to participate in the experiment than 10. The sound–only task was included in order to see if subjects would be generally able to perform the task without display, i.e. if the auditory feedback from the model might generally be a stand–alone alternative in tasks like the one here, whenever visual feedback is not available (e.g. for applications for visually impaired).

Task performance times with and without additional sonic feedback

Quite surprising after the preparatory considerations described above is the first main result of the performance experiment: *for all display sizes*, the average time needed to perform the task improves significantly with the auditory feedback from the model. Table 4.4 shows the average task times for individual subjects $(1, 2, \ldots, 10)$, the two groups $(1-5 \triangleq$ "with sound first", and $6-10)$ and the set of all subjects $(1-10)$ at the various display sizes, with and without sound. The two respective neighboring columns contain the relative difference, "no sound" to "with sound" (in %, δ) and the statistical p-value for the according set of measurements. p-values of (≤ 0.05) or near (≤ 0.1) statistical significance are highlighted in green. It can be seen that the average task time for the set of all subjects as well as for both subgroups improves (i.e. gets shorter) with the auditory feedback for all display sizes, corresponding to only positive δ-values (task time is longer without sound) in the last 3 lines (of table 4.4). These performance improvements, ranging from around 9% for the largest to around 60% for the smallest display, are always statistically significant for the whole set, while they reach statistical significance for the subgroups only for the smaller displays. Since significance is reached for the whole set of subjects, it can be expected that it would be found also for both subgroups, i.e. independently of the order of presentation with a sufficiently large set of measurements, using more subjects or more trials per subject.

Individual cases — of single subjects at a fixed display size — that contradict the general performance improvement, i.e. negative δ-values in table 4.4, are marked red. It is seen that all these (rather few) decrements of performance with sound are not statistically relevant, which justifies the expectation that these outliers are not systematic [25] and would tend to decrease in number and level with longer testing sessions. On the other hand, all individual differences

other hand. These points are discussed in detail in the respective paragraph.

[25] . . . i.e. not consistent signs of any regular mechanism of control behavior,

of, or close to, statistical significance (green p-values in the first 10 lines of table 4.4) are cases of improved performance with sound.

The slightly stronger performance improvement for group 2 at the largest

sub-ject(-s) no.	average task time (ms) at various display sizes, with (+) and without (−) sound, percentual difference (δ) and statistical significance (p)							
	scale factor 12				scale factor 4			
	+	−	δ (%)	p	+	−	δ (%)	p
1	6206	6828	10.0	0.282	7041	8437	19.8	0.276
2	4257	4295	0.9	0.933	4706	4370	−7.1	0.460
3	5795	7351	26.8	0.067	7009	9455	34.9	0.137
4	4767	5262	10.4	0.222	5009	6114	22.1	0.082
5	5908	5288	−10.5	0.433	6074	5480	−9.8	0.473
6	5478	5289	−3.4	0.701	4246	5700	34.2	0.004
7	4592	4599	0.1	0.987	4523	4685	3.6	0.741
8	5175	5516	6.6	0.554	6143	6430	4.7	0.732
9	5132	6846	33.4	0.037	6131	7241	18.1	0.298
10	4862	5475	12.6	0.244	5558	5650	1.7	0.902
1 – 5	5387	5805	7.8	0.203	5968	6771	13.5	0.135
2 – 6	5048	5545	9.9	0.063	5320	5941	11.7	0.086
1 – 10	5217	5675	8.8	0.031	5644	6356	12.6	0.029

	scale factor 2				scale factor 1			
	+	−	δ (%)	p	+	−	δ (%)	p
1	7313	8441	15.4	0.323	11004	NaN [24]	NaN	NaN
2	4539	5621	23.9	0.042	5710	12039	110.8	0.037
3	6782	8457	24.7	0.264	10718	18620	73.7	0.046
4	5599	6965	24.4	0.083	5907	9057	53.3	0.033
5	6551	7479	14.2	0.446	8361	17250	106.3	0.019
6	5631	8291	47.2	0.027	6430	6908	7.4	0.631
7	4994	5668	13.5	0.314	7013	11995	71.0	0.072
8	6615	7844	18.6	0.513	7888	6205	−21.3	0.155
9	8451	7793	−7.8	0.551	10713	29008	170.8	0.006
10	5446	6273	15.2	0.416	7972	7613	−4.5	0.830
1 – 5	6157	7392	20.1	0.015	8340	14242	70.8	0.000
2 – 6	6228	7174	15.2	0.095	8003	12346	54.3	0.018
1 – 10	6192	7283	17.6	0.004	8172	13188	61.4	0.000

Table 4.4: Average times needed to complete the "target matching"-task at the various display sizes, with and without sound. The additional columns contain the relative difference of the values δ, "without sound" to "with sound" in % and the statistical p-value for the two compared groups of measurements.

display size (scale factor 12) together with the smaller p-value, 0.0063 versus 0.203, might suggest that despite the training session of 2×10 trials we still have a slight learning effect that amplifies the positive difference of performance for group 2 and diminishes the effect for group 1. A direct comparison of the performances however, shows no significant difference between the results of the two groups, i.e. no significant influence of the order of presentation. Table 4.5 presents again the average task times for groups 1 and 2 in flipped orientation with the according p-values (well above 0.05). No particular reason has been found why group 2 overall solves the task faster than group 1, both with and without sound; the difference of averages (throughout, with and without sound) is however again not significant ($p = 0.159$).

	average task time		
	group1	group2	p
+ sound	5387	5048	0.209
– sound	5805	5545	0.425

Table 4.5: Average time needed by subjects 1 to 5 and 6 to 10 to complete the task, with and without sound, and the statistical significance.

A note has to be made concerning the smallest display, size 1. Prior to the experiment a clear improvement of performance was expected only for rather small displays, mainly size factor 1 and possibly 2. In fact, the difference of task times with and without sound is very high for this smallest size, as compared to the larger ones (an average of ca. 60% versus ca. $10 - 20\%$). At the same time the two outliers, subjects 8 and 10, and also subject 6 are in strong contrast to the rest of the test subjects, and to average results. Subject 8 performs better with display factor 1 without sound than for the next two bigger displays, 2 and 4 with and without sound; a similar statement holds for subject 6. These remarkable incidents may be due to an insufficiency in the display technique: at the chosen screen resolution [26] the line representing the rolling-track appeared not completely smooth on the screen, but small steps could be detected depending on the momentary angle of the line/track. One subject remarked that it was possible to recognize the exact horizontal position of the balancing-track by concentrating on keeping track of these steps in the display and that she used this phenomenon to steer the ball inside the target area even without clearly seeing it, rather by "intelligent guessing". This strategy that is not advantageous with larger displays, of course leads to a radical change in the perceptual or intellectual processes involved in solving the task. It may be responsible for the noted extreme outliers. As a consequence [27], in the following sections only such arguments are used, that can be sufficiently supported by the results for display factors 12, 4 and 2. In fact, in complete contrast to initial expectations,

[26] 1024×768, at higher screen resolutions, the graphical interface turned out to demand an unaffordable (in this context) amount of resources of computation.

[27] ... also in connection with the loss of measurements for subject 1 due to a technical problem at the first test run,

the measurements at the smallest display turned out to be unnecessary to prove any of the points of interest and might be left out. These values are therefore not "weighted" nor discussed specifically; they are however still displayed in all tables as they might be of informal interest. [28]

Mechanisms of performance improvement?

The results presented in the previous section (4.3.2) are strong arguments for the use of auditory display to support human–machine interaction in interfaces, environments or tasks as the presented one. In order to give more general specifications for the design of auditory display, consolidate the role and larger relevance of the on–hand example of sonic feedback and discuss supposable alternatives, it is important to look more deeply for mechanisms in subjects' behavior that lead to the noted performance increase with sound. Figure 4.3 depicts the situation.

I have argued earlier that sonic feedback in the real world is usually dynamic and continuous and it is here an important point to show that these respective qualities of the sound model are crucial from the standpoint of interaction performance and not "only" of esthetic value. [29] In fact, one might suspect that the average time to complete the task is shorter with sound, only because the controlling subject is additionally notified acoustically when the ball enters the target area, through the change in the rolling sound. It might be thinkable that subjects can simply react faster when the ball is entering the target area, and start earlier whatsoever stopping-manoeuvre. If this was the case, the dynamic quality of the sound feedback might appear as irrelevant for user performance; even more, no continuous sound feedback at all (at least outside the target area) might be necessary to gain the same auditory support of performance, just a short notification "ping" at the moment of entering the target area might have the same effect on the task times.

As a first step addressing the question just stated, the "*target reaching times*", i.e. the times for the virtual ball to reach (enter) the target area from its starting position, are surveyed, as depicted by the dash-dotted black line in figure 4.4. Average results under the different conditions of feedback are shown in table 4.6 in the format as known from table 4.4 ("with sound", "without", percentual difference, p-value). In the hypothetical case of irrelevance of the rolling sound outside the target area there should be no significant differences for the *target reaching times* with or without sound. Indeed it is seen that the average *target reaching time* for the set of all subjects does not significantly change with or without sound for any display size (last line of table 4.6). Especially for the two biggest display sizes, differences are very small, -0.2% resp. -2.4%. At first sight, this would support the hypothesis of potential irrelevance of the continuous rolling sound outside the target area. At a closer look how-

[28] . . . in particular for the design of possible future tests of more practice-oriented, quantitative focus.

[29] The previously reported tests of *sound recognition* (section 4.2) have already proved the potential of the rolling model in other respects. . . .

ever, strong (and sometimes strongly significant) differences of *target reaching time* are found for several individual cases at the three biggest display sizes, for single subjects and also for subgroup 1 (check for green p-values in table 4.6). Remarkably, these significant differences with sound are of opposite sign for different cases. [31] E.g., significantly shorter times with sound are found (positive δ-values) at display factor 12 (first main column) for subjects 9 and 10, or for subjects 4 and 6 at display factor 4 (second main column); opposite cases, i.e. negative δ-values, of significance are subject 3 at display factor 12, subject 7 at display factor 4 and subject 2 at display factor 2. The fact that for the largest display (factor 12) all significant (or close to significant) negative δ-values are in group 1 and all significant (or close to...) positive δ-values are found in group 2 raises the initial suspect of a pure training effect of some sort. The results for other displays however contradict this idea. Also, a comparison with table 4.4 shows that at display factor 12 subjects 3 and 9 both achieved remarkably (and

[31] ...and obviously basically cancel out in the average of the whole set.

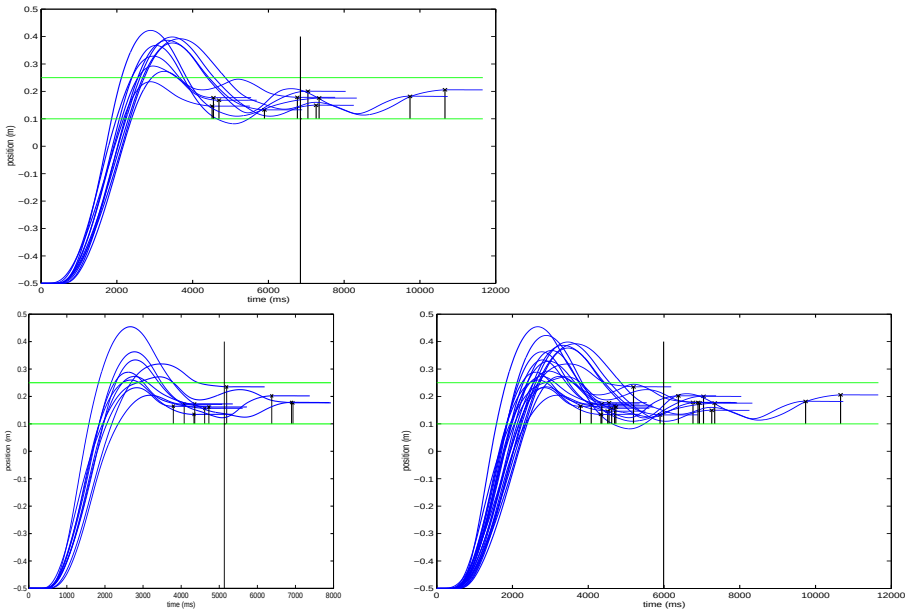


Figure 4.3: Trials of subject 9 at the largest display, without sound (10, above), with sound (10, below, left) and all 20 trails (below, right). A certain tendency of higher straightness/stability with sound can be seen, but clear, quantifiable mechanisms responsible for the improved average performance are not found from such overviews. In particular, the two groups of trials, with and without sound can not be separated in the (lower right) overall view. The following two sections deal with the extraction of various indexes from recorded data sets like these by statistical means.

significantly or almost significantly, $p = 0.067$) better performances with sound while their *target reaching times* are of opposite behavior. [32]

The *target reaching time* t_{target} is in each trial equivalent (exactly: antiproportional) to the ball's average velocity $\bar{v}$ before reaching the target area (see the black triangle in figure 4.4), via

$$\bar{v} = \frac{0.6\text{m}}{t_{\text{target}}} \tag{4.1}$$

Probably the next obvious value to observe for the ball moving towards the target area is its maximum velocity in that stretch (figure 4.4). The measurements, depicted in table 4.7, turn out to be of similar quality as the *target reaching times* and are thus discussed only very briefly: the average maximum velocity of the ball before reaching the target area for the set of all subjects is only slightly, and not significantly, different with or without sonic feedback (last line of table 4.7). Again, the distribution of differences for the largest display (see the opposite δ-values for group 1 and group 2), suggest the presence of a training effect, while some individual cases of significance, for display factors 4, 2 contradict this idea. In many cases, (average) maximal velocity and average velocity show a parallel behavior: for strong increases of average maximal velocity with sound — negative δ-values in table 4.7, see e.g. subjects 8 to 10 at display size 12 — we see increases of the average velocity, i.e. respective positive δ-values in table 4.6 (longer *target reaching times* without sound), and vice versa (at display size 12 e.g. subjects 2 and 3). For other subjects and displays,

[32] Both cases are significant in their average *target reaching times*, table 4.6.

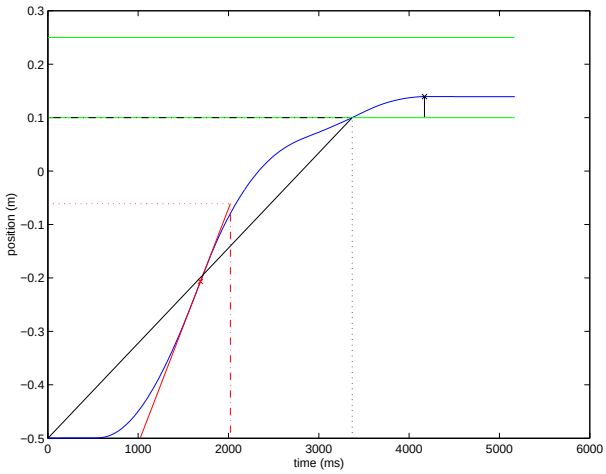


Figure 4.4: *Target reaching time* (ms, black dash-dotted line), average velocity and maximal velocity (m/s, length of the red dash-dotted line) for one example trial. Note that the horizontal width of the red triangle is 1s.

e.g. subject 7 at size 12 or subject 10 at size 4, average maximal and average velocity do not behave coherently: here the movement in average is obviously not just executed faster by a certain factor in one of the conditions (with or

subject(-s) no.	average *target reaching time* (ms) at various display sizes, with (+) and without (–) sound, percentual difference (δ) and statistical significance (p)							
	scale factor 12				scale factor 4			
	+	–	δ (%)	p	+	–	δ (%)	p
1	2528	2418	−4.3	0.394	2632	2339	−11.1	0.312
2	2462	2097	−14.8	0.077	2051	2163	5.5	0.600
3	3711	3010	−18.9	0.010	3823	3832	0.2	0.984
4	2612	2785	6.6	0.145	2385	2879	20.7	0.000
5	2633	2472	−6.1	0.160	2499	2421	−3.1	0.656
6	3415	3326	−2.6	0.640	2710	3197	18.0	0.003
7	3002	2883	−4.0	0.723	2752	2304	−16.3	0.017
8	2653	2993	12.8	0.070	3371	3037	−9.9	0.224
9	1962	2178	11.0	0.019	2093	2122	1.4	0.805
10	3089	3852	24.7	0.008	3389	2745	−19.0	0.019
1 – 5	2789	2556	−8.3	0.032	2678	2727	1.8	0.769
6 – 10	2824	3046	7.9	0.121	2863	2681	−6.4	0.145
1 – 10	2807	2801	−0.2	0.955	2770	2704	−2.4	0.520

	scale factor 2				scale factor 1			
	+	–	δ (%)	p	+	–	δ (%)	p
1	2196	2430	10.6	0.242	3267	NaN [30]	NaN	NaN
2	3217	2338	−27.3	0.004	2539	2813	10.8	0.511
3	4620	3900	−15.6	0.210	6070	9206	51.6	0.131
4	2400	2520	5.0	0.338	2433	2506	3.0	0.530
5	2534	2660	4.9	0.715	2391	2613	9.3	0.493
6	3280	3166	−3.5	0.699	3269	3159	−3.4	0.577
7	2822	2938	4.1	0.683	4048	3557	−12.1	0.214
8	3309	3284	−0.8	0.940	3329	3152	−5.3	0.594
9	2939	2443	−16.9	0.085	5202	5297	1.8	0.860
10	3378	3532	4.6	0.654	3943	3734	−5.3	0.667
1 – 5	2993	2769	−7.5	0.281	3340	4285	28.3	0.199
6 – 10	3146	3073	−2.3	0.613	3958	3780	−4.5	0.437
1 – 10	3070	2921	−4.8	0.241	3649	4004	9.7	0.363

Table 4.6: Average times for the controlled ball to reach the target area, at the various display sizes, with and without sound. Other columns contain the percentual difference “without sound” to “with” and the statistical p-value for the two according groups of measurements.

without sound) but changes qualitatively, i.e. in its general shape. [34]

All the present rather vague phenomena are indications that we see here traces also of deeper mechanisms of human control behavior influenced through

[34] It will be subject of the next section 4.3.2 to try and find, formalize and measure such qualitative changes of control movement.

sub-ject(-s) no.	average maximum velocity (ms) at various display sizes, with (+) and without (–) sound, percentual difference (δ) and statistical significance (p)							
	scale factor 12				scale factor 4			
	+	–	δ (%)	p	+	–	δ (%)	p
1	475	483	1.5	0.767	514	568	10.5	0.046
2	590	666	13.0	0.013	696	580	−16.6	0.033
3	297	366	23.1	0.028	310	309	−0.4	0.972
4	498	475	−4.7	0.315	525	479	−8.8	0.026
5	462	532	15.2	0.017	510	555	8.9	0.232
6	310	325	4.8	0.382	460	374	−18.7	0.000
7	461	397	−14.0	0.061	452	502	11.1	0.116
8	488	398	−18.6	0.038	351	399	13.9	0.207
9	617	527	−14.7	0.020	567	590	4.0	0.600
10	439	382	−13.0	0.044	447	460	2.9	0.489
1 – 5	465	504	8.6	0.078	511	498	−2.5	0.641
6 – 10	463	405	−12.4	0.009	455	465	2.1	0.625
1 – 10	464	455	−1.9	0.586	483	482	−0.3	0.931

	scale factor 2				scale factor 1			
	+	–	δ (%)	p	+	–	δ (%)	p
1	562	539	−4.1	0.418	371	NaN [33]	NaN	NaN
2	455	569	25.1	0.002	542	420	−22.6	0.020
3	276	306	11.0	0.115	257	257	−0.2	0.979
4	538	570	6.0	0.191	567	554	−2.4	0.669
5	569	573	0.7	0.910	614	577	−6.0	0.412
6	377	408	8.1	0.216	376	374	−0.6	0.927
7	494	404	−18.3	0.102	301	310	3.0	0.800
8	407	333	−18.1	0.008	402	379	−5.7	0.337
9	448	480	7.0	0.387	209	262	25.2	0.043
10	403	419	4.1	0.491	387	368	−5.0	0.512
1 – 5	480	511	6.5	0.204	470	452	−3.9	0.222
6 – 10	426	409	−4.0	0.318	335	339	1.0	0.843
1 – 10	453	460	1.6	0.654	403	389	−3.4	0.407

Table 4.7: Averages of the maximal velocity the ball reaches before entering the target area. The format of the other columns is as in previous tables.

sonic perception, not purely effects of training. In particular there are strong hints that the sonic feedback causes systematic differences of the movement of the controlled ball already before entering the target area. The initially possible suspect that the continuous sound feedback outside the target area is irrelevant appears now improbable. But the pure examination of *target reaching time* and average maximal velocity at this point does not reveal clear new explanations but instead raises even more questions. Further examination of the control movements is necessary to gain satisfying insights.

Differences of movement with and without sound

The first clear statements about an influence of the continuous sonic feedback on the control movements while solving the task can be made after extracting from the recorded trajectories the time at which the maximum velocity of the ball (before reaching the target area, as measured from the start of each trial) occurs. In figure 4.4 this is the temporal location of the red cross, referred to in the following as "*max.-velocity-time*". From table 4.8 holding the results (in the previously used format) it can be seen that in average over all subjects the ball reaches its maximum velocity earlier when the controlling subjects receive sonic feedback. This effect is present for all display sizes and always clearly significant, except for the smallest display. It is further seen that all individual cases (single subjects in table 4.8) of statistic significance [36] are supporting the rule, i.e. cases of earlier reached maximum velocity. — Subject 3 at display factor 2 is the only exception (out of 12 significant cases for the three largest displays). Vice versa, all other (than the latter) outliers, negative δ-values, marked red, are not significant. As is the case with the average *task performance times* (table 4.4) the clearer effect for group 2 and the two outliers at the largest display might suggest an influence of training that supports the auditory-based effect for group 2 and attenuates it for group 1. A t-test comparison of respective results of group 1 and 2, see table 4.9, shows a p-value close to statistical relevance in the "+sound" case.

Summing up "in plain words" the observed *max.-velocity times*, it can be said that subjects tend to accelerate the ball faster when they also *hear it*. More exactly, what I call "faster acceleration" is not simply a side effect of an overall faster movement since the maximum velocity itself was seen not to change significantly in average. Sonic feedback that would simply animate subjects to somehow move faster might not necessarily be advantageous since faster movements can also mean less precision and more error and thus more frequent and longer manoeuvres of correction. Instead it is seen here that the controlling subjects "save time" in the "right" phase of the movement, when accelerating the ball, without subsequently loosing control because of excessive maximum speed. It is seen that subjects use the additional information at their disposal in the sound to optimize their control movement. In particular, the phenomenon of more efficient acceleration shows that the continuous sonic feedback outside

[36] . . . or even all cases close to significance, green p-values,

the target area does have an influence on performance and can surely not be substituted by a short momentary notification signal. Naturally, more efficient acceleration in the beginning of the control task will lead to faster task completion if the gained temporal benefit is not lost later in the movement. The latter can be assumed, since the maximal velocity (in average) is not influenced

sub-ject(-s) no.	average max-vel.-time (ms) at display size, +/− sound, δ, p							
	12				4			
	+	−	δ (%)	p	+	−	δ (%)	p
1	1772	1927	8.8	0.055	1875	1840	−1.9	0.656
2	1446	1330	−8.0	0.375	1333	1516	13.7	0.091
3	2282	2210	−3.2	0.686	2453	2646	7.9	0.446
4	2306	2440	5.8	0.311	2058	2389	16.1	0.003
5	2103	2137	1.6	0.799	1819	1910	5.0	0.257
6	2340	2432	4.0	0.584	1870	2519	34.7	0.000
7	1548	1649	6.5	0.589	1441	1613	11.9	0.151
8	1987	2260	13.8	0.060	2510	2436	−3.0	0.754
9	1547	1854	19.8	0.001	1461	1659	13.5	0.023
10	1642	2184	33.0	0.020	1658	1841	11.1	0.110
1 – 5	1982	2009	1.4	0.767	1907	2060	8.0	0.120
6 – 10	1813	2076	14.5	0.005	1788	2013	12.6	0.024
1 – 10	1897	2042	7.6	0.027	1848	2037	10.2	0.007

	2				1			
	+	−	δ (%)	p	+	−	δ (%)	p
1	1667	1776	6.5	0.243	2496	NaN [35]	NaN	NaN
2	1641	1701	3.7	0.620	1604	2078	29.6	0.010
3	2411	2101	−12.9	0.038	2402	2807	16.8	0.036
4	2046	2242	9.6	0.012	2069	2279	10.2	0.057
5	1723	1929	11.9	0.094	1650	1970	19.5	0.014
6	2167	2323	7.2	0.322	2217	2501	12.8	0.142
7	1605	2045	27.4	0.044	2328	2366	1.6	0.798
8	1963	2063	5.1	0.606	2144	2090	−2.5	0.545
9	1756	1946	10.8	0.131	4037	3578	−11.4	0.330
10	1873	1937	3.4	0.538	1603	1815	13.3	0.291
1 – 5	1898	1950	2.7	0.454	2044	2284	11.7	0.001
6 – 10	1873	2063	10.1	0.014	2466	2470	0.2	0.980
1 – 10	1885	2006	6.4	0.020	2255	2387	5.9	0.162

Table 4.8: Averages of the time values at which the ball reaches its maximum velocity before entering the target area. Columns are of the same format as in previous tables.

through the sonic feedback, and it can thus be claimed that one, first reason for the better task-performance with sound has been found.

After the previous results of improved motion of acceleration with sonic feedback it is obvious to ask whether subjects also use the additional information in the rolling sound to optimize their movement while finally stopping the ball (or trying to...). Also, from the earlier (section 4.3.2) observation of unchanged average (overall) *target reaching times*, the presence of another systematic change in control movements while the ball is approaching the target can be deduced: if the improved task performance found its sole cause during the acceleration-phase, parallel significant changes in *target reaching times* as in the average task performance times should be found. With the aim of gaining more information about the stopping-movement, the velocity of the ball at the moment of entering the target area, referred to in the following as "*entry-velocity*" is extracted from the recorded trajectories. Again, average values for individual, the two groups and the set of all, subjects at the various display sizes with and without sound are shown, in table 4.10 (in the format known from previously observed indexes). It is seen that in average over all subjects the ball enters the target area slower when auditory feedback is present. This difference of average *entry-velocity* with and without sound is statistically significant for all display sizes but the largest. [38] Again significant differences are found also for several individual cases, all of which support the overall rule and are highly above average in their value. As for other previously discussed tendencies (shorter task times, earlier *max.-velocity times* with sound...) all outliers in table 4.10, i.e. all negative (red) δ-values, are clearly not statistically significant, according p-values are between 0.19 and 0.95. The observed lower *entry-velocity* with sound is another clear proof that the sound of the (virtual) rolling ball outside the target area has an influence on subjects' control behavior, since the phenomenon must be caused by a difference in control movement already before the moment of reaching the target. How, if at all, is this lower average *entry-velocity* related to other previously noted effects of sonic feedback, in particular to performance improvement, i.e. shorter average task times? Generally, it can be said that

[38] With the general difference of averages for the largest display not far from values of other display sizes, one statistically relevant individual case (see the following lines) and a overall p-value of 0.156 it is reasonable to believe that statistical significance could be reached for a larger set of subjects.

	max.-velocity time		
	group1	group2	p
+ sound	1982	1813	0.073
− sound	2009	2076	0.193

Table 4.9: Average times for which the ball reaches its maximum velocity for the two groups of subjects with opposite presentation order, "with sound" – "without" (group 1) and vice versa (group 2). Column 3 contains the statistical p-value for the respective sets of values.

for fast task performance it is desirable to stop the ball possibly shortly after it has entered the target area. To that end, any action aimed at stopping the ball should start already while approaching the target area. Starting from a fixed velocity outside the target area and assuming a given, fixed stopping-trajectory, task performance gets better, the closer to the left target boundary (after entering) the ball comes to rest; i.e. the slower the ball enters the target area,

sub-ject(-s) no.	average entry-velocity (mm/s) at display size, +/− sound, δ (%, p)							
	12				4			
	+	−	δ (%)	p	+	−	δ (%)	p
1	273	322	17.8	0.288	332	393	18.3	0.281
2	209	268	28.2	0.378	361	347	−3.8	0.862
3	123	183	48.7	0.196	170	199	17.2	0.528
4	411	371	−9.6	0.384	405	320	−21.0	0.197
5	297	374	25.8	0.118	272	289	6.3	0.816
6	149	212	42.3	0.019	159	236	48.5	0.043
7	135	181	33.9	0.259	127	282	122.3	0.004
8	276	259	−5.9	0.769	191	266	38.9	0.154
9	450	475	5.5	0.624	359	416	15.9	0.474
10	133	105	−21.1	0.341	138	231	66.7	0.056
1 – 5	263	304	15.6	0.136	308	310	0.5	0.958
6 – 10	229	246	7.8	0.562	195	286	46.8	0.001
1 – 10	246	275	12.0	0.156	251	298	18.5	0.032

	2				1			
	+	−	δ (%)	p	+	−	δ (%)	p
1	369	365	−0.9	0.956	259	NaN [37]	NaN	NaN
2	140	267	90.7	0.026	278	237	−14.8	0.519
3	110	180	63.8	0.040	86	111	28.8	0.351
4	417	489	17.3	0.317	418	490	17.3	0.277
5	293	324	10.4	0.658	278	311	11.8	0.590
6	185	227	22.1	0.347	163	252	54.2	0.050
7	179	206	15.3	0.571	79	208	161.7	0.008
8	191	199	4.5	0.849	165	201	21.9	0.405
9	210	346	64.9	0.042	123	174	41.6	0.198
10	152	134	−11.9	0.618	113	125	10.4	0.717
1 – 5	266	325	22.3	0.081	264	287	8.8	0.575
6 – 10	183	222	21.3	0.088	129	192	49.1	0.001
1 – 10	225	274	21.9	0.022	196	234	19.3	0.043

Table 4.10: Average velocity of the ball at the moment of entering the target area. The format is identical to previous tables.

and vice versa. Figures 4.5 (a) and (b) serve to explain this idea. Stopping (or trying to...) the ball shortly after entering the target area, very close to the boundary, on the other hand increases the risk of "stopping too early" and thus having to correct, in this way loosing time; an example of such a case is shown in figure 4.5(c). From the average values in table 4.10 (lowest line), it

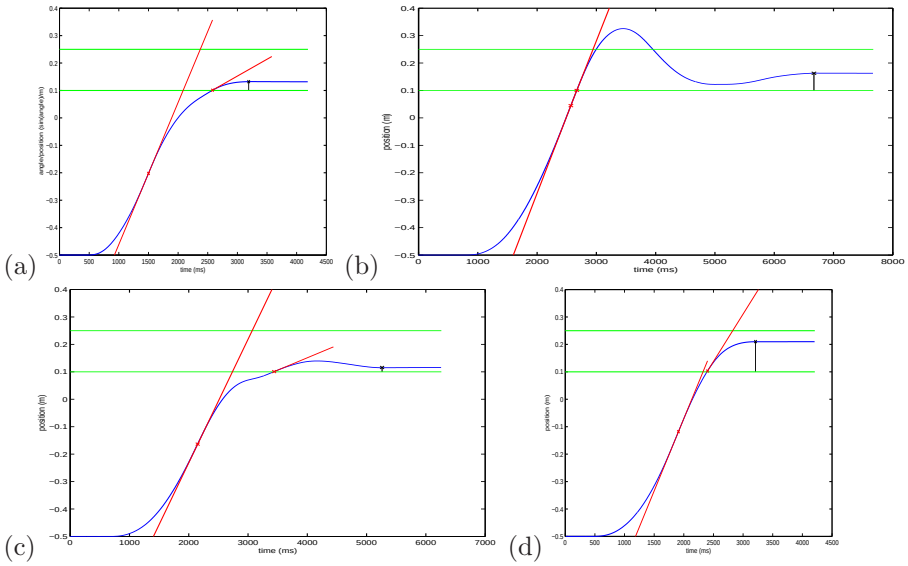


Figure 4.5: Examples for the possible (tendential) connection of *entry-velocity* and task performance. (a) depicts a somewhat ideal control movement (with accordingly very fast task performance; subject 2, largest display, with sound): the ball, that reaches its maximum velocity rather early, is slowed down "just in time" to stop shortly after entering the target area. In a less optimal example, (b) (subject 4, largest display, no sound), the ball enters the target area with maximum velocity (that is reached only shortly before) and the controlling subject subsequently does not directly succeed to stop within the target area. The contrary extreme (subject 4, display size 2, no sound) is shown under (c): the stopping movement is too strong/early, the ball has to be accelerated again to reach the target area and time is again lost in the final correction phase. An exception to the rule (as compared to (a)) is shown under (d): here (subject 4, largest display, with sound) the ball enters the target area closer to its maximal velocity because the stopping-manoeuvre is started rather late, but the latter is very efficient resulting in an equally good task time.

has to be assumed that generally subjects exploit the additional information available from the rolling sound to optimize their stopping-manoeuvres in the sense just stated. With sonic feedback, in average subjects appear to be able of stopping the ball earlier without increased risk of "stopping too early". This is the first notion suggested by the parallel phenomena of improved task perfor-

mance and slower *entry velocities* with auditory feedback. The latter idea can also serve to explain why improved task performance overall is not connected to shorter *target reaching times*, as asked in the beginning of this paragraph: earlier stopping-motion with the ball coming to rest earlier after entering the target area can also increase the time span of reaching the target area. For example is the target area in figures 4.5 (a) and (b) reached at approximately the same time (around 2.5s, despite the faster acceleration in (a), due to the earlier stopping-phase. Such an effect would counteract the "headstart"-effect of more efficient acceleration with auditory feedback.

Summing up the last considerations the following picture is gained of how the movement of control and thus of the ball during the task changes when auditory feedback is added:

- In average, subjects use the additional information about the reaction/motion of the controlled ball conveyed through the sound, to optimize their control movements such that the ball **1**. accelerates faster in the beginning and reaches its maximum velocity earlier and **2**. slows down earlier, indicated through lower average *entry-velocity* and stops earlier after having entered the target area. As a side-effect, the *target reaching time* stays basically unchanged in average, while task performance times improve with sound.

The overview of results of subject 6 at display size 4, figure 4.6, (a) with and (b) without sonic feedback serves well to exemplify the previous principle.

Of course this picture is to be seen as a model for the average tendency of control movements, not as an exhaustive strict rule. At the beginning of the preceding section 4.3.2 in figure 4.3 I have already noted an overall tendency of movement-trajectories to appear more straight or "controlled" with the presence of auditory feedback. It has to be assumed that the stopping-manoeuvre (out- and inside the target area) in average does not only start earlier but gets also more efficient, i.e. shorter as a fruit of additional sonic information. This is what was seen to happen with the initial acceleration-phase and a hypothesized absence of the parallel mechanism for the stopping-phase appears unreasonable. A shorter stopping-phase on the other hand might again lead to higher *entry-velocities*, i.e. attenuate or annihilate the effect stated above. Figure 4.5 (d) may serve as an example for this notion.

If individual task times are compared, *max.-velocity times* and *entry-velocities*, i.e. single elements of tables 4.4, 4.8 and 4.10 (averages for single subjects at specific display sizes) it can be seen that in all cases of improved task performance, i.e. positive δ-values in table 4.4, at least one of the two effects noticed lately, faster acceleration, i.e. positive δ in table 4.8, or lower entry-velocity, i.e. positive δ in table 4.10 is found. — The only exception to this rule, subject 10 at display size 2, might be a case of exceptionally efficient, i.e. short, stopping-manoeuvres, following the preceding consideration. In some cases, lower *entry-velocity* with sound accompanies a lower maximum velocity (positive δ-values in table 4.7) which may explain performance outliers (negative δ-values in table 4.4) as subject 5 at display sizes 12 and 4 or subject 9 at size

2. *Entry velocity* and *max.-velocity time* are after all seen to be useful indexes to prove and measure qualities of control movement.

Purely auditory feedback

All 10 subjects were able to perform the task with purely auditory feedback only. The overall average task time with sound only was slightly better (i.e. shorter) than at the smallest display without sound. This result is seen as informal or preliminary since the sound–only case was always presented as the last, the order of presentation (for these two cases) was not varied, although it appears very improbable that there is still an effect of training present towards the end of the testing session (after ca. 1h). More important, I consider a direct comparison purely of task times with sound and with a small display as rather uninteresting: it is clear that the task becomes unsolvable for displays below a certain size, so necessarily task times will become longer for sufficiently small displays than for the sound–only condition — as long as the task is solvable purely with sound. The latter showed to be the case and that is in fact the important lesson from this part of the test. It would surely be interesting to plan and execute more thoroughly, comparative measurements with purely auditory and purely visual feedback in future tests. A deeper analysis of control movements in those two cases might further support the general insight from the tests, that subjects perceive and exploit different information through the

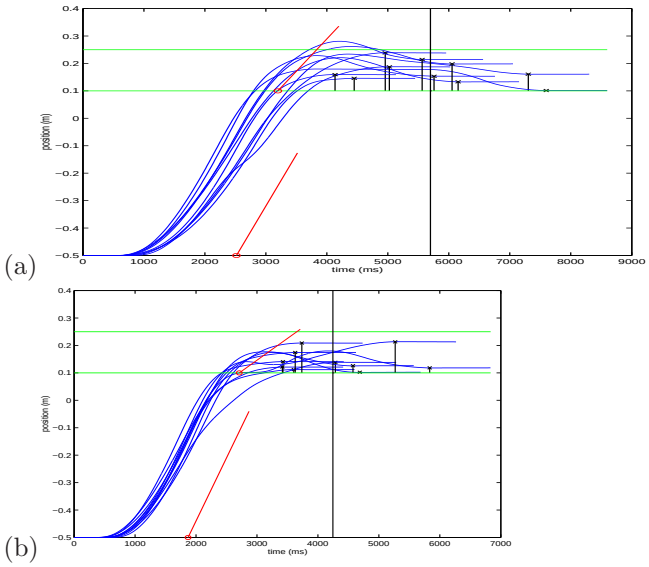


Figure 4.6: Overview over the ten trials of subject 6 at the second largest display, above without (a), below with sound (b). With sonic feedback, in average the maximum velocity is reached earlier, the ball enters the target area with lower average velocity and the task is completed faster.

two different sensory channels — visually mainly position, velocity auditorilly — and possibly reveal more details.

4.3.3 Outlook

Naturally I do not claim to have captured and perfectly formalized with the indexes introduced in the last section, all possible mechanisms of movement/-perception behind the discovered phenomenon of higher performance with auditory display in the above test task — rather scratched the surface. Another strategy of approaching the phenomenon of varying performance would be the development of a model of the performing subject e.g. in terms of a differential equation or related transfer function. This approach is practiced in the field of robotics where models of the *human operator* e.g. in teleoperation tasks are constructed [16][44]. It would surely be a promising scope, to integrate results from the measurements into such a model. The derivation of a *human operator model* accounting also for auditory perception indeed seems new terrain. Maybe the ideas, measurements and qualitative pictures given above could be a basis for further developments in the direction. However, the direct intuitive exploitation of continuous auditory information in human control behavior appears not to have been proven or measured at all before, which gives uniqueness to the work described in the last chapter.

Interesting is the confrontation of spontaneous remarks of test subjects during the experiment with measured results. Subject 1 stated her suspect or subjective feeling, that for the largest display size, solving the task does not get easier with auditory feedback. The according results however show an improvement of her performance under these conditions (table 4.4, line 1, column 1). Also the two finally used indexes (tables 4.8 and 4.10) in this case (line 1, column 1) behave according to the main picture. Another subject remarked that solving the task is "much harder" without sound, since the surface is "more slippy": of course the objective behavior of the virtual ball de–facto does not change. Remarkable in the same sense was the reaction of yet another subject, who was "sure", i.e. convinced, that the virtual ball inside the target area with its rougher surface profile reacts different than outside; this is again really not the case. The statement might be followed further and inspire tests concerning the auditory perception of surface roughness, connecting to the respective work of Lederman [43]. All these remarks are hints that the sound model might also be used for the benefit of the *subject-centered* [39] qualities of interactive environments, multimodal systems or interfaces, in the sense shortly sketched in the beginning of this chapter. After the last cited user comment above, the expressive strengths of the model even promise its use for the examination as well as exploitation of phenomena of *sensory substitution*, i.e. the provocation of sensory impressions that are usually connected to another perceptual channel than the stimulated one. Finally, it can be seen that human subjects are generally not aware of their measured intuitive perception and exploitation of

[39] I pick up the term as introduced at the beginning of the chapter in section 4.1.

auditory information. It thus has to be assumed that the respective results here would not be detected by means of more "straightforward" evaluation methods, e.g. listening tests plus questions or scaling- and sorting-tests.

4.4 Central conclusions concerning the rolling model and auditory display

To "close the cycle" towards the initial impulses of the sound design work that has been assessed (at an example application) here, towards its bases in the *ecological viewpoint* of psychoacoustics as well as new demands in human–computer interaction, the central points of the preceding evaluation tests and consequences, concerning the presented sound model as well as auditory display in general and the specific approach here, shall be summarized:

- The sound generated by the sound model showed (section 4.2) to provoke a very strong spontaneous connotation of an "archetypical" scenario, that of rolling. This means that it conveys or supports with high reliability the idea of a specific way of interaction (namely rolling) between two objects that are also quite specific in their attributes and behavior: a round (basically, not too edgy) solid object is moving in continuous contact with low friction (no sliding or scratching), rotating, on a smooth (to a certain degree) surface, with "well-behaved" velocity and direction... Using the realtime implementation of the model, its potential to spontaneously steer a users expectations on the behavior of a system, purely acoustically or in direct synchronization with other perceptual modes in an interface, is finally seen. The sound modeling work (chapter 2) is thus seen to be successful and the specific approach of using a hybrid sound design architecture suitable.

- In section 4.2 it has been found that the ecologically expressive yet abstracted (i.e. not necessarily perfectly realistic) model can in some cases outperform real sounds in terms of clear identification. This demonstrates the efficiency of the idea of *cartoonification* (section 1.2) as a dynamic auditory pendant to graphical cartoon icons, and introduces and exemplifies its application in realtime *reactive* implementations.

- As already noted in the first argument, the model is shown to be directly and successfully embedable in a larger control interface including and synchronizing other modes of interaction, here visual and gestural. In this connection with a familiar, convincing metaphor its potential to steer a users approach towards a system and define his/her way of (inter)action is seen to be particularly strong. The use is intuitive in the sense of requiring virtually no explanation or training, as opposed to abstract sonifications. In the performance test of section 4.3.1, subjects were given only the goal of the task, no explanations whatsoever of how to achieve it; the handling of the device proved to be selfexplanatory.

- I have suggested in chapter 1 that sound can convey continuous ecological information perceived by the human auditory system and that this mechanism might and should be used for the benefit of human–computer interaction. Previous works of psychoacoustic research have demonstrated the auditory perception of ecological attributes through questionnaires, labeling- or sorting tasks. The performance experiment described in section 4.3 on the other hand, proves the perception and exploitation of acoustic information through measurements of control movements, without affecting (instead: retaining and underlining) the intuitive, unaware nature of the process. Not only the perception of continuous acoustic information is shown but simultaneously its direct exploitation in optimized control behavior. To my knowledge, the performance test and its results are unique in these respects.

- Besides proving the superior potential of the rolling model over the common use of short sound signals of notification/warning, a possible way to use reactive auditory display in interaction tasks, supporting or replacing graphical display, is demonstrated. Clear concrete performance measures, task completion times, are given.

- The sound engine of the *Ballancer* relies only on the rolling model and does not use complex and costly spatialization. Position information in detail, i.e. apart from the distinction of the target area from the surrounding rolling-plane, is expressed only by stereo-panning — i.e. very roughly. As a consequence, the optimization of control movements and task performance with sound, shows subjects' ability to perceive and exploit the information of velocity contained in the sound; position information is here available visually with higher precision. Further on, a clear optimization of control and performance (contrary to previous expectations) also while using a very large display (complete computer screen), can be noted. This demonstrates that enhanced continuous auditory display as used here, can not only compensate for restrictions of graphical display of practical reasons (such as small display size) but open generally new ways of information transfer: It has been shown here that velocity information can be perceived and exploited from sound, as had been proposed in the beginning of section 4.3, while it can not directly be perceived [40] visually.

- Since the task showed to be generally solvable also with purely auditory feedback, the sound model is seen to be potentially useful in similar situations also by itself. This aspect is interesting e.g. for applications for visually impaired and could surely be strengthened through the inclusion of state–of–the–art algorithms of spatialization.

[40] I here refer to the remarks concerning the perception of velocity at the beginning of this section (4.3) and about the linkage of perceptual channels and perceivable information as discussed in chapter 1 (section 1.1); I am not discussing a possible relation to the psychological concept of *direct perception*.

Conclusion

As computers shall be embedded, "disappear", become pervasive, ubiquitous, wearable..., in other words merge with our familiar surroundings and allow "natural" interaction, we need to provide human–computer interfaces with a sonic channel that is adequate to the omnipresence of sound and the significance and immense potential of human auditory perception. To this end we need knowledge about human auditory perception and strategies and techniques to generate informative sound. Work and progress in the two directions are strongly dependent and interconnected in many ways and this thesis has contributed to both aspects, although its initial, central scope lies in the second area, in providing tools for auditory display and the use of sound in human–computer interaction. In fact, those new achievements of the thesis that are of direct psychoacoustic interest, are results of the evaluation experiments (chapter 4) performed with, and dedicated to, one of the sound models developed in the first major part (chapter 2).

The contribution of the last chapter to general psychoacoustic knowledge is however not completely surprising, not simply a side effect of the efforts to push further sound in human–computer interaction, as the background chapter (1) has already argued towards the necessary connection of the potential information to be conveyed and the employed *channel(s)* of human perception. The psychoacoustic notion of *everyday listening* has been introduced in its main points seen as relevant here, and in its relation and partial contrast to traditional psychoacoustic theories and tools. It has been proposed that mechanisms of *everyday listening* can not only be used to the benefit of human–computer interaction (as proposed and done before) but that the employment of this traditionally unused (in human–computer interfaces) perceptual (sub)channel may open new qualities of interaction, i.e. allow the conveyance of information that can so far not be transmitted to a user, not just of "more information"; this claim has finally been proved in the last chapter of the thesis (4). *Continuous* reactivity of sonic feedback, as omnipresent in "the real world", has been pointed out as an important factor towards the latter goal, in continuation of the existing (and now already classic, see W. Gaver) pioneering implementations of everyday-like sounds in human–computer interfaces. The term "sound models" has been chosen to stress on continuous, reactive quality, expanding the older concept of *auditory icons* which also shares the main notion of *cartoonification*.

Concrete realizations of the general scopes have been developed and imple-

mented in the "technical" chapter 2, concentrating on the significant class of ecological sounds originating from scenarios of impact-based contacts of solid objects. This work forms a contribution to the area of sound synthesis, as it extends existing physical models, expands their traditional scopes and practical use and delivers several sonic scenarios (bouncing, breaking sounds) that have not been achieved previously, or with less degree of detail and realism (rolling). The question of *modular* realtime implementation, that is a common issue in physics-based sound synthesis, has also been attacked successfully; the achieved modularization is an important point as it allows an easy expansion of the sound model catalog presented here [41]. Further on, the main aspects of the concrete development of the sound models, in its psychoacoustic and technical approach, have been conceptualized for sound design, thus suggesting a more general value of the sound modeling work, as expandable examples rather than an arbitrary isolated list of implementations. It has been proposed and demonstrated that physics-based models can be used under more differentiated principles than for simulation as a goal in itself, with the aim of *cartoonification* and under integration of signal-based methods. This idea has been summarized as a *hybrid* hierarchical architecture for sound design.

The interactive potential of the developed sound models, synchronized with other modes such as vision and gesture, and the practicability, e.g. of the modal parameters, has been demonstrated at some examples of interactive devices reported in chapter 3. Of these, the *Ballancer* is of particular value since it exemplifies the solid realization of a "sound–friendly" control metaphor, and allows in its robustness and simplicity to demonstrate the value of enhanced sonic feedback in addition to, or even substitution of, graphical display.

Through the evaluation of the rolling model and *Ballancer* in chapter 4 the initial claims of the thesis and the suitability of the sound design approach and work in reaching these scopes, have finally been verified — the circle is closed. The found strong connotation of the synthesized sound with rolling, in confrontation with recorded "real" sounds, concretely demonstrates the idea and value of *cartoonification*. The perception of ecological information (momentary velocity) from sound and its exploitation in optimized control movements of users — spontaneously, without conscious explanation or training — has been proven. This result is of relevance beyond demonstrating the success and usefulness of the sound model (of rolling), as it uncovers and provides evidence for a phenomenon, the direct, steered gestural reaction on continuous auditorily perceived information, that appears "natural" but has never been demonstrated before. Besides, this performance test has introduced an unconventional, "indirect" strategy of assessing a perceptual mechanism. Since the test does not rely on questions (about the guessed velocity), subjects are not biased by additional implications and conscious responses; an effect of the auditory perception is revealed without and beyond subjects' awareness.

[41] The example of the friction *module* has been mentioned repeatedly, that uses code and the implementational structure presented in chapter 2.

"Dort am Klavier, lauschte sie mir,
und als mein Spiel begann, hielt sie den Atem an."

There at the piano, she listened to me,
and as my playing began, she held her breath.

Rammstein [42]

[42] Rammstein: *Sehnsucht, Klavier* 1997

Sommario

Affinché i computer vengano integrati nel nostro ambiente naturale, sparendo come unità distinte e diventando pervasivi, indossabili o ubiqui, abbiamo bisogno di fornire all'interazione uomo–macchina un canale sonoro che sia di complessità confrontabile a quello che è l'onnipresenza del suono nel mondo reale. Non possiamo più permetterci di sprecare le immense capacità della percezione uditiva umana in interfacce uomo macchina che ci forzano a guardare costantemente uno schermo e ad usare i canali acustici solo per segnali di notifica occasionali statici e ripetitivi. Gli *auditory displays* devono diventare reattivi/dinamici, continui e intuitivamente informativi come i suoni che ci circondano e accompagnano le nostre azioni negli ambienti quotidiani.

Per tanto tempo l'uso e la conoscenza del suono sono stati limitati dall'attenzione unilaterale della psicoacustica tradizionale sugli attributi astratti del suono, come altezza, intensità o brillantezza, e dalle restrizioni di metodi classici di generazione del suono (e.g. sintesi sottrattiva o FM), che sono controllati in termini di parametri come frequenze e ampiezze e si basano su di essi. Lo standard attuale del suono nei sistemi computerizzati, ovvero la riproduzione di campioni fissati di suono, che si può vedere come la prima reazione alle restrizioni descritte, non è soddisfacente per via del suo carattere statico, ripetitivo, nè reattivo e nè dinamico. I precedenti ostacoli per l'apertura di nuovi raffinati "auditory displays" comunque cominciano a dissolversi attraverso sviluppi recenti sia nella psicoacustica che nella generazione del suono.

La scuola ecologica di psicoacustica mette in evidenza il significato della percezione uditiva umana come (forse il primo) trasporto di informazione sui processi nei nostri ambienti quotidiani, e che nell'ascolto di ogni giorno noi percepiamo le *fonti del suono* piuttosto che gli attributi musicali o dei segnali. Il crescente numero di lavori nel campo pone le basi per i rispettivi sforzi di una sintesi del suono *ecologicamente espessiva*. Questi non devono necessariamente risultare in imitazioni di suoni reali provenienti dagli ambienti quotidiani. Spesso, ad un estremo realismo è preferibile una caricatura mediante esasperazione di alcuni importanti attributi ecologici di un complesso scenario familiare (nel senso di icone grafiche o *auditory icons*) al prezzo di altri attributi considerati di minore interesse. Tuttavia, malgrado un certo lavoro in questo senso, c'è ancora spazio libero per ulteriori sviluppi nel campo della generazione del suono; la formulazione e l'esplorazione di un'impostazione sistematica e più generale verso la realizzazione e l'utilizzazione di idee di *espressione uditiva ecologica*

e "*cartoonification*" è un obiettivo considerevole.

In particolare non è stato stabilito un collegamento più profondo e sistematico tra le varie tecniche di sintesi del suono esistenti, comprese le più nuove, e l'approccio psicoacustico sopra menzionato, considerando anche aspetti di utilizzazione e implementazione. Questo può anche riflettere i ruoli tipicamente assegnati all'espressione uditiva e visiva. Il suono è generalmente riconosciuto nella sua enorme rilevanza come il mezzo della lingua e della musica. Ma, mentre ogni bambino sa disegnare gli "smileys" o altre icone del fumetto, è ancora necessario un orientamento di base per avvicinarsi ad un "sound design" espressivo in senso ecologico, e un'efficiente progettazione del suono. Dal punto di vista ecologico è di alto interesse una tendenza piuttosto recente nella generazione del suono, nota come "physical modelling" e basata su descrizioni fisico–matematiche di sistemi meccanici che emettono suono piuttosto che su proprietà di segnali (da generare). Comunque la maggior parte dei lavori nel campo riguarda la simulazione possibilmente realistica di singoli e unici sistemi fisici, principalmente strumenti musicali. Le implementazioni risultanti sono tendenzialmente troppo complesse nel controllo e nel calcolo per essere usate come parte di un'interfaccia uomo macchina. [43]

Solo recentemente ha iniziato a svilupparsi un collegamento più profondo e dedicato, che congiunge l'esperienza della sintesi del suono basata sulla fisica e le speculazioni della "*psicoacustica ecologica*" [44].

Argomento generale della tesi

Il lavoro presentato propone un percorso per superare o migliorare la situazione attualmente decentrata e poco qualificante del display uditivo. Daremo strumenti e valideremo un approccio al sound design per fornire all'interazione uomo–macchina un utilizzo migliore e innovativo del canale uditivo, neccessario all'indispensabile e incessante percezione umana dell'informazione acustica di contesti "naturali". Introducendo il concetto di *everyday listening*, basandoci su *espressioni ecologiche* piuttosto che su segnali di tipo astratto, arriveremo alla comprensibilità intuitiva, ovvero la comprensione spontanea e la cattura delle reazioni dell'utente senza esplicazione o training preliminare. Useremo il termine "sound model" per riferirci agli algoritmi di generazione del suono che abbiamo sviluppato, i quali incorporano un comportamento sonoro dinamico (complesso) piuttosto che (collezioni di) campioni fissati e a sè stanti. Questo sforzo verso una reattività continua è un'importante prosecuzione dell'idea delle *auditory icons*, le quali condividono l'aspetto caricaturale dell'espressione ecologica *cartonificata*. Il secondo aspetto importante del nostro concetto di sound

[43] Naturalmente il canale uditivo dell'interfaccia di un sistema non può utilizzare la stessa quantità di risorse computazionali di uno strumento interamente dedicato al suono, quale è uno strumento musicale elettronico, e generalmente sono altamente specializzate e piuttosto inflessibili nel loro potenziale sonoro.

[44] . . . e.g. nel corso del progetto di ricerca europeo "*The Sounding Object*" [67] su cui ha lavorato l'autore di questa tesi, e che ha fortemente ispirato e influenzato il lavoro qui presentato.

design è l'applicazione di tecniche allo stato dell'arte di sintesi sonora, più precisamente l'utilizzo di modelli fisici. Contraramente ad altri lavori di modellazione fisica, andremo alla ricerca e deriveremo astrazioni quando queste siano utili per flessibilità, basso costo computazionale e implementativo, e chiarezza di espressione. Durante questo processo la conoscenza e i punti di forza delle tecniche convenzionali di sintesi del suono non vengono ignorate, ma piuttosto parimenti sfruttate, risultando infine in un'architettura ibrida che combina tecniche basate sulla fisica e tecniche basate sul segnale all'interno di strutture centrate sull'aspetto percettivo.

Come conseguenza del loro comportamento dinamico e della loro reattività in tempo reale, i nostri modelli per il suono possono essere naturalmente combinati e sincronizzati con altri canali percettivi, attraverso il display grafico o una periferica di ingresso gestuale. L'inquadramento robusto all'interno di chiare e possibilmente note metafore globali per l'interazione con (o il controllo di) un sistema, può consolidare ulteriormente la comprensione a livello intuitivo. L'adeguatezza e il successo del nostro lavoro concettuale e di sviluppo è provata dall'esempio del modello di rotolamento e dal *Ballancer*, un "gioco" interattivo audio-visio–tattile (uno degli esempi di realizzazione multimodale). Questi test di valutazione inoltre confermano e quantificano il miglioramento della perfomance dell'utilizzatore attraverso l'utilizzo di un feedback acustico reattivo e informativo continuo, come avviene correntemente nelle nostre azioni nelle situazioni di ogni giorno (viceversa mancante negli attuali ambienti d'interazione uomo–macchina). Il capitolo sulla valutazione si distingue per particolare originalità in quanto risultati di tale chiarezza, oppure associati a un'applicazione quale quella qui presentata, non sono mai stati dimostrati in altra sede.

Bibliography

[1] 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.

[2] F. Avanzini. *Computational issues in physically-based sound models*. PhD thesis, Università degli Studi di Padova, 2001.

[3] 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.

[4] 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.

[5] Federico Avanzini and Davide Rocchesso. Impact. In Davide Rocchesso and Federico Fontana, editors, *The Sounding Object*, pages 125–129. Mondo Estremo, Firenze, Italy, 2003.

[6] S. Barrass. *Auditory Information Design*. PhD thesis, Australian National University, 1997.

[7] D. R. Begault. *3-D Sound for Virtual Reality and Multimedia*. Academic Press, 1994.

[8] J. Bensoam, N. Misdariis, C. Vergez, and R. Causs. Finite element method for sound and structural vibration: musical application with modalys sound synthesis program based in modal representation. In *SCI: Systemics, Cybernetics & Informatics*, Orlando, Florida, 2001.

[9] Joel Bensoam. A reciprocal variational approach to the two–body frictionless contact problem in elastodynamics. *International Journal for numerical methods in Engineering*, 2002.

[10] J. Blauert, H. Lehnert, J. Sahrhage, and H. Strauss. An interactive virtual-environment generator for psychoacoustic research. i: Architecture and implementation. *Acta Acoustica*, 86:94–102, 2000.

[11] S. Bly. *Sound and computer information presentation.* PhD thesis, University of California, Davis, 1982.

[12] Giapaolo Borin, Giovanni De Poli, and Davide Rocchesso. Elimination of delay-free loops in discrete-time models of nonlinear acoustic systems. *IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING*, 8(5):597–605, September 2000.

[13] Giapaolo Borin, Giovanni De Poli, and Augusto Sarti. Algorithms and structures for synthesis using physical models. *Computer Music Journal*, 16(4):30–42, 1992.

[14] R. Bresin, S. Dahl, M. Marshall, M. Rath, and B. Moynihan. Controlling the virtual bodhran - the vodhran. In *Stockholm Music Acoustics Conference (SMAC) 2003*, Stockholm, Sweden, 2003.

[15] S. A. Brewster. Non–speech auditory output. In J. Jacko and A. Sears, editors, *The Human-Computer Interaction Handbook*, pages 220–239. Lawrence Erlbaum Publishers, 2002.

[16] Thurston L. Brooks, editor. *Telerobot Response Requirements.* STX Robotics, Lanham, 1972.

[17] W. Buxton. Using our ears: an introduction to the use of nonspeech audio cues. In E. J. Farrel, editor, *Extracting Meaning from Complex Data: Processing, Display, Interaction*, pages 124–127. Proceedings of SPIE, Vol 1259, 1990.

[18] 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.

[19] P. R. Cook. *Real Sound Synthesis for Interactive Applications.* A. K. Peters Ltd., 2002.

[20] I. Daubechies and S. Maes. A nonlinear squeezing of the continuous wavelet transform based on auditory nerve models. In A. Aldroubi and M. Unser, editors, *Wavelets in Medicine and Biology*, pages 527–546. CRC Press, 1996.

[21] H. Q. Dinh, N. Walker, C. Song, A. Kobayashi, and L. F. Hodges. Evaluating the importance of multi-sensory input on memory and the sense of presence in virtual environments. In *Proceedings IEEE Virtual Reality*, Houston/Texas, USA, March 13–17 1999.

[22] Yoshinori Dobashi, Tsuyoshi Yamamoto, and Tomoyuki Nishita. Real-time rendering of aerodynamic sound using sound textures based on computational fluid dynamics. In *Siggraph 2003*, San Diego, 2003.

[23] Kenneth P. Fishkin, Anuj Gujar, Beverly L. Harrison, Thomas P. Moran, and Roy Want. Embodied user interfaces for really direct manipulation. *Communications of the ACM*, 43(9):75–80, 2000.

[24] P. M. Fitts. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, 47:381–391, 1954.

[25] D. Fjellström. Investigating motor memory and tactility for recall of abstract entities. Master's thesis, Umeå University, 2002.

[26] F. Fontana, L. Ottaviani, M. Rath, and D. Rocchesso. Recognition of ellipsoids from acoustic cues. In *Proc. Conf. on Digital Audio Effects*, pages 160–164, Limerick, December 2001.

[27] D. J. Freed. Auditory correlates of perceived mallet hardness for a set of recorded percussive events. *J. Acoust. Soc. Am.*, 87(1):311–322, January 1990.

[28] Zhi-Fang Fu and Jimin He. *ModalAnalysis*. Harcourt, 2001.

[29] W. W. Gaver. *Everyday listening and auditory icons*. PhD thesis, University of California, San Diego, 1988.

[30] W. W. Gaver. How Do We Hear in the World? Explorations in Ecological Acoustics. *Ecological Psychology*, 5(4):285–313, Apr. 1993.

[31] William W. Gaver. Using and creating auditory icons. In G. Kremer, editor, *Auditory Display: Sonification, Audification, and Auditory Interfaces*, pages 417–446. Addison-Wesley, 1994.

[32] Bruno L. Giordano. Everyday listening: an annotated bibliography. In Davide Rocchesso and Federico Fontana, editors, *The Sounding Object*, pages 1–14. Mondo Estremo, Firenze, Italy, 2003.

[33] D. E. Hall. Piano string excitation VI: Nonlinear modeling. *J. of the Acoustical Society of America*, 92:95–105, July 1992.

[34] C. Hendrix and W. Barfield. The sense of presence within auditory virtual environments. *Presence: Teleoperators and Virtual Environment*, 5(3):290–301, 1996.

[35] M. M. J. Houben, A. Kohlrausch, and D. J. Hermes. Auditory cues determining the perception of size and speed of rolling balls. In *ICAD01*, pages 105–110, Espoo, Finland, 2001.

[36] M. M. J. Houben and C. N. J. Stoelinga. Some temporal aspects of rolling sounds. In *Journée design sonore à Paris*, Paris, France, 2002. < http://confs.loa.espci.fr/DS2002/ >.

[37] Y. Hua. Parameter estimation of exponentially damped sinusoids using higher order statistics and matrix pencil. *IEEE Transactions on Signal Processing*, 39(7):1691–1692, July 1991.

[38] E. Hutchins, J. Hollan, and D. Norman. Direct manipulation interfaces. In D. A. Norman and S. W. Draper, editors, *User Centered System Design: New Perspectives in Human-Computer Interaction*. Lawrence Erlbaum Associates, 1986.

[39] Julius O. Smith III. Physical modeling synthesis update. *Computer Music Journal*, 20(2):44–56, 1996.

[40] 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.

[41] A. J. Kunkler-Peck and M. T. Turvey. Hearing shape. *J. of Experimental Psychology: Human Perception and Performance*, 26(1):279–294, 2000.

[42] S. Lakatos, S. McAdams, and R. Caussé. The representation of auditory source characteristics: simple geometric form. *Perception & Psychophysics*, 59(8):1180–1190, 1997.

[43] S. J. Lederman. Auditory texture perception. *Perception*, 8:93–103, 1979.

[44] Sukhan Lee and Hahk Sung Lee. A kinesthetically coupled teleoperation: Its modelling and control. *IEEE*, 1991.

[45] R. A. Lutfi. Auditory detection of hollowness. *J. Acoust. Soc. Am.*, 110(2), August 2001.

[46] R. A. Lutfi and E. L. Oh. Auditory discrimination of material changes in a struck-clamped bar. *J. Acoust. Soc. Am.*, 102(6):3647–3656, December 1997.

[47] I. S. MacKenzie. *Fitts' law as a performance model in human-computer interaction*. PhD thesis, University of Toronto, 1991.

[48] Thomas Mann. *Doktor Faustus*. Fischer Verlag, Frankfurt am Main, Germany, 1988(1947).

[49] 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.

[50] B. C. J. Moore. *An introduction to the Psychology Of Hearing*. Academic Press, 4th edition, 1997.

[51] C. Müller-Tomfelde, N. A. Streitz, and R. Steinmetz. Sounds@work - auditory displays for interaction in cooperative and hybrid environments. In C. Stephanidis and J. Jacko, editors, *Human-Computer Interaction: Theory and Practice (Part II)*, pages 751–755. Lawrence Erlbaum Publishers, 2003.

[52] Public enemy: *It takes a nation of millions to hold us back*. Def Jam Recordings, 1988.

[53] Roy D. Patterson. The sound of a sinusoid: Spectral models. *J. Acoust. Soc. Am.*, 96(3):1409–1418, September 1994.

[54] Roy D. Patterson. The sound of a sinusoid: Time-interval models. *J. Acoust. Soc. Am.*, 96(3):1419–1428, September 1994.

[55] Pure data, *pd*. < http://www.pure-data.org >.

[56] The *Radio Baton*. < http://emfinstitute.emf.org/exhibits/radiobaton.html >.

[57] D. Rocchesso. Acoustic cues for 3-d shape information. In *ICAD (International Conference on Auditory Display)*, pages 180–183, Espoo, Finland, July 2001.

[58] Davide Rocchesso and Federico Avanzini. Discrete-time-equations. In Davide Rocchesso and Federico Fontana, editors, *The Sounding Object*, pages 124–125. Mondo Estremo, Firenze, Italy, 2003.

[59] Davide Rocchesso and Federico Avanzini. Friction. In Davide Rocchesso and Federico Fontana, editors, *The Sounding Object*, pages 129–136. Mondo Estremo, Firenze, Italy, 2003.

[60] Davide Rocchesso and Federico Avanzini. Impact. In Davide Rocchesso and Federico Fontana, editors, *The Sounding Object*, pages 125–129. Mondo Estremo, Firenze, Italy, 2003.

[61] Davide Rocchesso and Federico Fontana, editors. *The Sounding Object*. Mondo Estremo, Firenze, Italy, 2003. < http://www.soundobject.org >.

[62] 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.

[63] Sabine Rückert. Die Erhörte. *Die Zeit (Leben)*, 34:53, August 12th 2004.

[64] M. R. Schroeder. New results concerning monaural phse sesitivity. *J. Acoust. Soc. Am.*, 31:1579(abs), 1959.

[65] Rammstein: *Sehnsucht*. Universal, 1997.

[66] author's website. < http://www.sci.univr.it/ rath >.

[67] *The Sounding Object (SOb).* European research project (IST-25287, < http//:www.soundobject.org >) as part of the *Disappearing Computer (DC)* proactive initiative (< http://www.disappearing-computer.org/ >).

[68] C. N. J. Stoelinga, D. J. Hermes, A. Hirschberg, and A. J. M. Houtsma. Temporal aspects of rolling sounds: A smooth ball approaching the edge of a plate. *Acta Acoustica*, 89:809–817, 2003.

[69] A. Stulov. Hysteretic model of the grand piano hammer felt. *J. of the Acoustical Society of America*, 97(4):2577–2585, Apr 1995.

[70] K. van del Doel. *Sound Synthesis for Virtual Reality and Computer Games.* PhD thesis, University of British Columbia, 1998.

[71] K. van del Doel, P. G. Kry, and D. K. Pai. Foleyautomatic: Physically-based sound effects for interactive simulation and animation. In *Proc. ACM Siggraph 2001*, Los Angeles, Aug. 2001.

[72] K. van den Doel, D.K. Pai, T. Adam, L. Kortchmar, and K. Pichora-Fuller. Measurements of perceptual quality of contact sound models. In *ICAD (International Conference on Auditory Display)*, Kyoto, Japan, July 2002.

[73] Kees van den Doel and Dinesh K. Pai. The sounds of physical shapes. *Presence*, 7(4):382–395, August 1998.

[74] N. J. Vanderveer. *Ecological Acoustics: Human perception of environmental sounds.* PhD thesis, Georgia Institute of Technology, 1979. Dissertation Abstracts International, 40, 4543B. (University Microfilms No. 80-04-002).

[75] H. von Helmholtz. *Die Lehre von den Tonempfindungen.* Olms Verlag, Hildesheim, Germany, 1968(1862).

[76] H. L. F. von Helmholtz. *On the Sensations of Tone.* Longmans, Green and Co., London, UK, 1912. English translation of 4th edition by A. J. Ellis.

[77] W. H. Warren and R. R. Verbrugge. Auditory perception of breaking and bouncing events: a case study in ecological acoustics. *Journal of Experimental Psychology: Human Perception and Performance*, 10(5):704–712, 1984.

[78] R. Wildes and W. Richards. Recovering material properties from sound. In W. Richards, editor, *Natural Computation*, pages 356–363. MIT Press, Cambridge, MA, 1988.