Federico Fontana

Physics-based models for the acoustic representation of space in virtual environments

Ph.D. Thesis

16 Marzo 2003

Università degli Studi di Verona
Dipartimento di Informatica

Université Bordeaux 1 - U.F.R. Math-Info
INRIA - équipe "Signes"

Advisor:
prof. Davide Rocchesso

Series N°: **TD-06-03**

Università di Verona
Dipartimento di Informatica
Strada le Grazie 15, 37134 Verona
Italy

a Rossella,
alla sua disponibilità e comprensione

Contents

Preface to the Final Edition

This edition of the thesis comes out in an aim to unify the style of the durable scientific documents produced in my Department, to collect them into series. It does not contain technical novelties.

To this effort I have personally contributed under the guidance of the Director of the PhD, Andrea Masini, by preparing the LaTeX template that is now used in this series. I thank Andrea for supervising this activity, and for its contribution to the final look of the final version of my thesis. I also thank Isabella Mastroeni who has provided hints to improve the template.

Federico Fontana
Train IC 604 "Cattaneo" from Venice to Verona, 8 Feb 2006

Preface

[...] dissertò di dominanti e colori del suono, di un centro neutro permanente
dal quale col suono apportava o levava energia a qualunque sostanza.
Ma c'era un problema: il generatore funzionava solo quando c'era lui,
e Keely era geloso di quel segreto che accordava la materia allo spirito.
Può egli aver scoperto qualcosa ed essere incapace di esprimerla?
[Nikola Tesla, osservazione sul Globe Motor di J. E. Worrell Keely. 1889.]

[...] he spoke of dominants and colors of sound, of a permanent neutral centre
which he gave or took away energy from, with the sound to whatever substance.
But there was a problem: the generator worked only with him,
and Keely was jealous of that secret which gave form to spirit.
May he have discovered something and not be able to express it?
[Nikola Tesla, observation on the Globe Motor made by J. E. Worrell Keely. 1889.]

The first time I went to a conference of computer music, in 1995, I was amazed by the many panels presenting papers on acoustics, psychoacoustics, audio signal processing, applied computer science and audio engineering, which addressed musical performance, the art of composition, the specification of ideal listening and the ways to achieve that ideal. This musical emphasis was demonstrated by the conference's social event, normally including an artistic performance such as a concert of contemporary music: I later realized that this was a *must* in most of those conferences.

In the science of sound there is a clear link between research and art that is less common than other disciplines.

The study of the everyday listening experience and its systematic[1] inclusion as a branch of research in sound is quite recent. Perhaps this new branch has not yet been totally accepted by some research communities in the field. Obviously, during most of their lifetime humans use the hearing system to acquire information from the external world, but they spend little time appreciating the acoustical subtleties of state of the art technology such as high-end sound synthesis and reproduction equipment. Likewise, most of the researchers working in other disciplines focus on

[1] and not necessarily artistic, as it has been happening since long time ago: think, for instance, of Luigi Russolo's *intonarumori* [46, 132].

solving problems that, directly or not, concern practical matters; rarely do they deal with the artistic derivations of their research.

Do we choose to do study in the everyday listening spectrum only because our methods of investigation cannot deal with the artistic level of the hearing experience? At least two arguments can be given to nullify that suspicion. First, most everyday sounds are so familiar to the listener that they can be hardly reproduced by a model accurately enough to cheat the hearing system, if not the subject. Second, art must not be used to mystify fair research results, as it happens sometimes in this field.

To put it plainly, we would say that models that simulate everyday listening, including those reproducing the spatial location of a sound source, those synthesizing ecological sounds, and more generally those adding some sense of presence to a virtual scenario, are prone to more criticism and severe evaluation compared, for instance, to models for the synthesis of hybrid sounds or for the automatic generation of scores.

On the other hand most of those actually doing research in sound are (or were) sound practitioners or expert music listeners, if not musicians, composers or professionals in the musical field. Those people ought to consider a different research point of view, where quality is evaluated by means of psychophysical tests, and musical sounds are substituted by ecological events such as bounces, footsteps, rolling, crushing and breaking sounds.

This is not an easy perspective to adopt, especially for those having a musical background. It is opinion of the author that accepting this new perspective will be a key step in helping the science of sound to achieve a prominent position among the most respected scientific disciplines.

Federico Fontana
Pordenone, 2 Feb 2003

1

Introduction

Qualora si voglia comprendere il senso delle trasformazioni reali
dell'attività di progettazione (nel XX secolo) sarà necessario
costruire una nuova storia del lavoro intellettuale e della sua lenta
trasformazione in puro lavoro tecnico (in 'lavoro astratto', appunto).
[. . .] Inutile piangere su un dato di fatto: l'ideologia si è mutata
in realtà, anche se il sogno romantico di intellettuali che
si proponevano di guidare il destino dell'universo produttivo
è rimasto, logicamente, nella sfera sovrastrutturale dell'utopia.
[Manfredo Tafuri, *La sfera e il labirinto: avanguardie e architettura da Piranesi agli anni '70.* 1980.]

Whenever you want to understand the sense of the real transformations
of the design activity (in the 20th century) it would be necessary
to create a new history of the intellect and of its slow transformation
into a pure technical job (into an 'abstract job', in fact).
[. . .] It's unnecessary to cry over a matter of fact: ideology muted into reality,
even if the romantic dream of intellectuals, who proposed to lead the productive
universe's destiny, still remains, logically, in the overall sphere of utopia.
[Manfredo Tafuri, La sfera e il labirinto: avanguardie e architettura da Piranesi agli anni '70. *1980.]*

This work deals with the simulation of virtual *acoustic spaces* using physics-based models. The acoustic space is what we perceive about space using our auditory system. The physical nature of the models means that they will present spatial attributes (such as, for example, shape and size) as a salient feature of their structure, in a way that space will be directly represented and manipulated by means of them.

We want those models to be provided with input channels, where *anechoic* sounds can be injected, and output channels, where *spatialized* sounds can be picked up. We also expect that the spatial features of those models can be modified even during the simulation (namely, runtime) through the manipulation of proper *tuning parameters*, without affecting the effectiveness of the simulation. Once these requirements are satisfied, we will use the models as sound synthesizers and/or

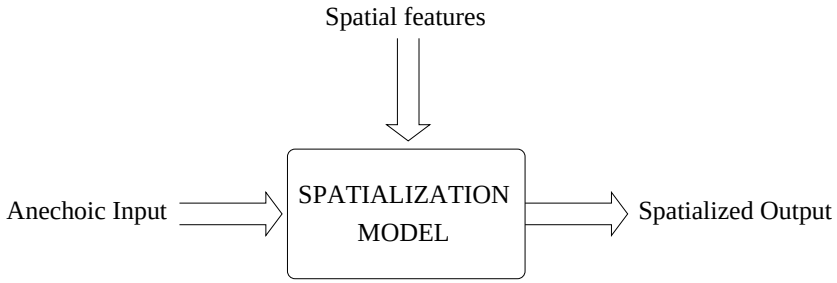

Fig. 1.1. Spatialization model

sound processors to provide anechoic sounds with spatial attributes, in the way expressed by Figure 1.1.

Finally, we will use our models as reference prototypes for the design of alternative, simplified representations of the auditory scene when such representations appear to be more versatile than the original prototypes, especially for application purposes.

As we will see in more detail in Chapter 2, sound spatialization (primarily concerned with the representation of *reverberant* environments) has an important place in the history of audio. Especially during past decades there has been some difficulty in adding convincing spatial attributes in real-time to sound. This has led to a large variety of solutions which provided satisfactory reverberation in the majority of cases.

On the other hand, the spatial attributes provided by those solutions were often qualitative. In other words, most of the geometrical and structural figures characterizing the spatialization domain (such as the shape of a resonator, the size of an enclosure or the absorption properties of walls) have rarely been considered worth rendering.

Indeed, the qualitative approach is supported by psychoacoustics. Since humans can only roughly evaluate structural quantities from spatialized sounds, any effort to design models that can be tuned in the quantitative parameters would possibly result in applications whose performance does not justify their technical complexity. Nevertheless, this consideration applies especially to musical listening contexts, where the salient features of the auditory scene and the acoustic quality of the listening environment relate in a complicated and unpredictable way to the structural parameters of the physical environment.

Conversely, let us consider the everyday listening experience. There the auditory system is engaged in different tasks, mainly recognition and evaluation. One of those, for example, is distance evaluation. For the aforementioned reasons, the traditional literature about spatialization by means of reverberation has not too much to share with those tasks.

There is a branch of the field of Human-Computer Interaction (HCI), called *auditory display*, where the subjective evaluation of a listening scenario is seriously considered. From the literature concerned with the psychophysics of human auditory localization we learn that there is a gap between the precision of the physical laws that calculate the sound pressure level in the vicinity of the listener's ears,

caused by the sonic emission from a sound source located somewhere in a listening environment, and the unpredictability of the listener's judgment about the spatial attributes of the same environment. We say that the *perceived* spatial attributes differ from their corresponding objective quantities.

Due to its nature, a physics-based model for sound spatialization can be easily accommodated to simulate a listening environment whose attributes are conveyed to the listener in the form of auditory spatial *cues*. Using that model we can reproduce those cues, manipulating their properties by directly controlling the physical quantities that characterize the listening environment.

The quality of the model will be evaluated through listening tests. We cannot keep subjective differences under control, but this is not a major concern for this work, since we are not interested in tailoring models to optimize subjective performance. Conversely, we want to validate those models by comparing the subjects' overall subjective evaluations with results found in alternative experiments conducted in *real* listening environments.

In one case concerning the auditory recognition of three-dimensional cavities, we have used the model also as a reference listening environment, since we did not find psychophysical experiments in the literature to be taken as a reference counterpart.

Following a general introduction to sound spatialization in Chapter 2, Chapter 3 deals with some preliminary theoretical facts about the *Waveguide Mesh*, the numerical scheme which we will extensively use to give an objective representation of auditory space. This chapter is interesting especially for those who are familiar with this scheme, although its comprehension highlights the advantages and actual limits of the Waveguide Mesh as a spatialization tool.

In Chapter 4 we propose an example of *sonification* in between sound synthesis and sound spatialization. In this chapter, we will examine the motivations for synthesizing sounds from models of ideal resonant cavities, having shapes ranging from the sphere to the cube. We will also show the method we have followed to simulate the aforementioned cavities, and discuss the results obtained by the corresponding simulations, along with their interpretation.

Chapter 5 describes a virtual listening environment for the synthesis of distance cues realized using the Waveguide Mesh. Due to its *tubular* design, this environment can be easily constructed. Results of its evaluation are discussed basing on subjective ratings on the effectiveness of the cues provided by that virtual environment. Those results indicate that physics-based modeling can play a role in the design of spatialization tools devoted especially to the presentation of information in multimodal human-computer interfaces.

Chapter 6 addresses the question of presenting auditory data. This chapter emphasizes the problems that arise when spatialized sounds finally have to be presented to the listener and the possible consequences of a wrong presentation, even using state-of-the-art reproduction systems. Although the chapter does not break any new ground in this specific field, it nevertheless surveys some facts related to headphone presentation, and defines a framework in which our spatialization model can be successfully employed. Part of the chapter is devoted to presenting a novel structure of *parametric equalizers*, useful for canceling the most audible

artifacts coming from distortions caused by the audio chain and peculiar defects of the listening environment where the audio reproduction system is located.

Finally, it is our conviction that a model which simulates the acoustic space must process sounds coming from the everyday listening experience. For this reason, Chapter 7 of this work only briefly addresses the specification and synthesis of a certain type of ecological sound that we will call, with some generality, *crumpling* sound. Those sounds result from a superposition of individual bouncing events according to certain stochastic laws, in such a way that varying the bouncing parameters and/or the stochastic law translates into a change of the "nature" of the resulting sounds. Hence, they can represent the crushing of a can, or a single footstep. The results obtained in this chapter will be used as an autonomous set of anechoic samples, to feed the spatialization models.

Whenever a careful treatment of at least part of a topic mentioned in any chapter can be found, references will be given. In particular, articles already written or co-authored by the same author of this work, covering the subjects proposed herein, will be reported in the second part of the thesis. This choice has been made with the purpose of avoiding rearrangements and condensations of arguments that have already been presented, and to save the reader from too much detail on any particular topic.

Nevertheless, an effort has been made to keep the first part of the thesis self-contained, in such a way that the general sense of the overall work can be understood from that part. Whenever needed the reader can refer to the second part.

Several audio examples have been produced using the tools developed during this work. Actually, they are meaningful only in an ongoing research process. For this reason, here we include the address of the author's web site, where proper links to those examples can be found. This address is profs.sci.univr.it/∼fontana/

Finally, a convention in the use of pronouns should be pointed here. We will refer to subjects acting in a listening context (such as talkers) using feminine pronouns. On the other hand, listeners and, more in general, machine users will be identified with masculine pronouns.

2

Physics-Based Spatialization

Già a metà del Settecento Julien Offroy de La Mettrie descriveva
ne *L'Homme-Machine* la trasformazione del dualismo uomo-natura
in quello uomo-macchina, anche se l'assunzione del modello macchina
come spiegazione della realtà fisica risale alla metà del XVII secolo.
[Vittorio Gregotti, *Architettura, tecnica, finalità*. 2002.]

By the middle of the Eighteenth century Julien Offroy de La Mettrie described in the book "L'Homme-Machine" the transformation of the man-nature dualism into man-machine dualism, even if the assumption of the machine model as explanation of the physical reality, dates back to the middle of the Seventeenth century.
[Vittorio Gregotti, Architettura, tecnica, finalità. *2002.]*

The only place to hear sounds that are unaffected by spatialization cues is the inside of a good anechoic chamber. In that listening situation, sound pressure waves either reach the listening point or fade against the walls of the chamber. More precisely, both the sound source and the listener should be as *transparent* to sound waves as possible. Ideally, both of them should be point-wise and have the same directivity and sensitivity along all directions. In all other cases, sounds are affected by some kind of spatialization.

2.1 Spatial cues

Consider the case of a group of persons talking inside an anechoic chamber: each speaker, in addition to her personal voice, has a direction-dependent emission efficiency (or *directivity*), and diffracts the emitted acoustic waves with her head and body, so that, for example, she can be heard by listeners even if she does not face them. Each listener partially absorbs, partially reflects, and, in his turn, diffracts those waves before they reach the listeners' ears. Objects which are present in the scene behave as reflectors and diffractors as well, and in addition can transmit acoustic waves. Moreover, speakers are also listeners, so that their presence in the scene determines changes similar to those caused by the listeners. This peculiar example, referred to an anechoic chamber, shows that there are very few situations

in which sounds are not spatialized, so that in almost all listening contexts humans can in principle extract spatial cues from sounds, that give information about the relative position, orientation and distance of the sound source, and about the listening environment.

Despite the existence of spatial information in the sounds, humans are not able to exploit that information in a way that they can extract a precise scene description from it, but grossly. In fact, spatial attributes contained in unfamiliar sounds in general do not translate into reliable cues, such as loudness or pitch, that enable humans to perform precise detections [171]. An exception comes from the detection of the angular position of a sound source, that can be assessed even with good precision for certain sources. This happens due to *binaural listening*, and due to the existence of the *precedence effect*, but not due to the spatial information that is inherently contained in the sound as it approaches the listener's proximity.

To stay in the example of the talkers in the anechoic chamber, a listener can normally detect the angular position of the speaker with good precision and, to some extent, recognize her distance and her orientation inside the chamber given that her voice is familiar to him, i.e., once some prior training has been conducted inside the anechoic environment. He can also detect the presence in the chamber of a large object if it has a sufficient reflectivity.

In the following of this chapter we will briefly survey human auditory spatial perception. In fact, a basic knowledge of spatial hearing cannot be avoided by the reader interested in this work. Since excellent research has been done in the subject, the reader will be often referred to the literature.

We will pay more attention in outlining the psychophysics of distance perception. In fact, one goal of this research is to provide models for the synthesis of virtual attributes of distance. Thus, a comprehension of the mechanisms enabling distance perception is considered to be fundamental for the reader who wishes to capture the meaning of the following chapters, and the overall sense of this work.

2.1.1 Binaural cues

Binaural cues give humans a powerful way to detect the angular position of a sound source. In fact, the hearing system is capable of processing the time differences of individual low-frequency components contained in a sound, that reach the two ears at different arrival times due to the interaural distance. These cues are known in the literature as Interaural Time Differences (ITD), and enable humans to position the sound source into a geometrical locus (the so-called *cone of confusion*) that is consistent with those time differences [15].

Moreover, the hearing system can also detect the intensity differences of individual medium- and high-frequency components contained in a sound, that reach the two ears at different amplitudes due to the masking effect caused by the head (also called *head shadow*). These cues, which are binaural as well, are known in the literature as Interaural Intensity Differences (IID) or Interaural Amplitude Differences (IAD) [15].

ITD and IID are complemented by the *precedence effect*, by means of which binaural differences contained in signal onsets are taken in major consideration by the hearing system. As a consequence, the angular position of a sound source

detected during the attack of a sound is retained for a few tens of milliseconds, regardless of the positional information contained in the signal immediately following the onset [15, 124].

Head diffraction gives humans also the possibility, to some extent, to recognize the distance of a sound source located in the near-field (that is, within about 1 m) [22, 145]. The synthesis of binaural cues providing near-field distance information is not object of study in this work, for reasons that will be made clear in the following.

Binaural cues can be effectively reproduced by models that calculate the ITD and IID according to the angular position of the sound source [21]. Those models in general leave some degree of approximation in the reproduction of such cues: in fact, binaural cues are generally considered to be quite *robust* against subjective evaluation.

IID can also be modeled to reproduce the distance of a nearby source [41].

2.1.2 Monaural cues

Before reaching the ear canal, sound waves are processed by the human body. Torso, shoulders, and in particular the two (supposed identical) ear *pinnas*, add spatial cues to a sound according to the subjective anthropometric characteristics of each individual. Those cues do not bring independent information to each one of the two ears, hence they are called *monaural*[1].

Moreover, during their path from the acoustic source to the listener's proximity, sounds are modified by the listening environment. As we will see later in more detail the environment is responsible for important modifications in the sound, which are grouped here together with the monaural effects of the human body. Although from a physical viewpoint such modifications result in independent sounds reaching the two ears, the binaural differences existing in those "environmental" attributes do not convey binaural cues as well.

Monaural cues enable humans to recognize the elevation and distance of a sound source, along with other features which are not object of study in this work [166]. As a rule of thumb, cues caused by the listener's body play a major role in the detection of the elevation, whereas cues added by the environment are mainly responsible for distance evaluation. Despite this, environmental modifications mainly consisting in sound reflections over the floor have been reported to convey elevation cues [68]. On the other hand, interferences caused by the head against the direct sound originating from nearby sources have been hypothesized to be responsible for monaural distance cues by Blauert [15].

It is straightforward to notice that monaural cues are less reliable than binaural cues. In fact, their existence in a sound can be detected by the hearing system as soon as the source (i.e., free of monaural cues) sound is familiar to the listener [15]. For the same reason, dynamic changes in the sound source or the listener's positions have positive effects in the detection of elevation and distance, since during those changes (that turn into variation in the corresponding cues) the hearing system

[1] Though, inequalities between the left and right side of the body, in particular pinna inequalities, have been reported to convey binaural cues accounting for elevation [141].

can familiarize with the original sound, and, thus, better identify the cues added by the listener's body and the listening environment.

Monaural cues which are added by the body have a quite subtle dependency on the listener's anthropometric characteristics: for this reason they are highly subjective, and much more difficult to reproduce than binaural cues. Nevertheless, models exist aiming at evoking a sense of elevation [3, 8]. In general, these models tend to work better if they can handle dynamic variations of the elevation parameters once they have been informed about the listener's head position [57].

For their importance in this work, monaural cues added by the listening environment are dealt with in the following section.

2.1.3 Distance cues

Distance detection involves the use of many (possibly all) monaural cues that are present in a sound. In the simple case of one still sound source playing in a listening environment in the medium- or far-field, our hearing system recognizes an *auditory event* to happen somewhere in the environment.

It is likely that the localization of the auditory event, in particular its distance from the listener, is somehow related with the actual position of the sound source. In the case of distance perception, the gap existing between the position of the sound source and the localization of the auditory event is influenced by several factors, and in many cases it cannot be easily explained or satisfactorily motivated. Judgments on distance are influenced by visual cues as well [144].

Also, the so-called distance *localization blur* [15] has been discovered to be non-negligible, main consequence of the subjectivity of the psychophysical scales on which subjects rely during the evaluation.

In general, psychophysical tests investigating distance perception are difficult to design and set up, sometimes contradicting, and (especially in the case of prior investigations) often prone to criticism on the methodological approach, and suspicion concerning the quality of the equipment used for the tests [15].

Experiments aiming at studying the effects of one singular cue, although keeping a high degree of control over the test, led sometimes to results having low practical consequences. Nevertheless, those experiments have shown that three monaural cues are important for distance perception [167]:

loudness – this cue plays a fundamental role especially in the open space, and once the sound source is familiar to the listener (the last fact is frequently expressed in the literature saying that loudness is a *relative* distance cue). In an ideal open space or anechoic environment, each doubling of the source/listener distance causes the sound pressure at the listener's ear to decrease by 6 dB [102]. Hence, if the source emits a familiar sound (particularly speech), then the hearing system is capable of locating the auditory event based on the loudness resulting from the corresponding sound pressure at the ear entrance.
Experiments where the pressure at the ear canal was kept under control have shown that the logarithmic distance of the auditory event is mapped by the sound pressure value at the ear, measured in dB, according to an approximatively linear psychophysical law—at least within specific ranges and for some

sound sources. Moreover, the localization blur approximately follows the same law [15, 167].

The steepness and offset of this law is subjective to a large extent. In addition to that, there is evidence that the type of sound emitted by the sound source has a noticeable influence on that steepness and on other scaling factors, depending on the sound nature (e.g., sinusoids vs. clicks) and familiarity of the listener to that sound (e.g., shouted vs. whispered speech) [56].

Loudness provides also an *absolute* cue (i.e., independent of other factors) if dynamic loudness variations are taken into account. In fact, in this case the subject can in principle exploit the aforementioned 6-dB law (in practice one approximation of it). Experiments investigating the effect of dynamic loudness variations on distance detection show a better performance of subjects who were enabled to use that cue [66].

direct-to-reverberant energy ratio – once the listening experience takes place in the inside of a reverberant enclosure, each sound source is perceived not only by the direct signal, but together with all the acoustic reflections reaching the ears after the direct sound (see Section 2.2). More precisely, the perceived reverberant energy is a property of the enclosure and the source and listener positions inside it, and, hence, it conveys absolute positional information to the listener as long as it has experience of that particular listening environment.

Although loudness should yield reliable distance cues, but relative in the case of unfamiliar sounds, nevertheless several experiments showed that subjects perform good distance detections when they listen to reverberant sounds, and not only in the case of unfamiliar sounds and/or familiar reverberant environments [66, 108, 169]. That evidence indicates that listeners can take advantage of such absolute cues.

This conclusion seems to be reasonable, considering that anechoic environments are almost ever experienced during everyday listening (even open spaces provide reflective floors, except when they are covered with soft snow or other strongly damping materials [32]). One more reason supporting this conclusion may be that enclosures having similar volume and geometry, normally experienced in everyday life, exhibit overall reverberant properties that are often quite indistinguishable by listeners.

spectrum – air exhibits absorption properties over sounds, which become perceptively non-negligible as long as acoustic waves travel in this medium for more than approximately 15 m [15]. The corresponding cues, normally perceived as low-pass filtering increasing with distance, should be noticeable both in the open-space case and inside enclosures, by long-trajectory reflected sounds.

Their importance in distance detection is debated. In fact, sources located farther than 15 m, i.e., beyond the so-called *acoustical horizon* proposed by Von Békésy [162], hardly result into definite auditory localizations, but in the context of particular scenarios. On the other hand, low-passed sound reflections usually appear in environments where reverberation provides a rich and composite acoustic complement to the direct sound even neglecting air absorption.

Most of the experiments on distance perception show that subjects tend to overestimate the distance of sources in the near-field, and underestimate the distance of sources in the far-field. This has suggested the existence of a perceptual

factor, called *specific distance tendency* [97], that relocates the auditory distance to a "default" value as long as distance cues are removed from the sound. Zahorik [167] has proposed this value to be around 1.5 m, in good agreement with the results coming from listening tests in which he was able to provide sounds virtually free of distance cues.

Zahorik has also proposed to look at the various disagreements existing between different experiments as effects of a general rule, according to which subjects must arrive to a unitary auditory scene, provided with a coherent notion of distance, by averaging and "trading" among several types of cues coming from different sound sources, whose "quality" inside the listening environment varies to a broad and partially unpredictable extent.

2.1.4 The HRTF model

The auditory cues we have seen in § 2.1.1 and § 2.1.2 can be all included in a stereophonic signal, obtained by exciting the listener's near-field using particular sounds[2], and simultaneously measuring the individual responses at the left and right ear canal entrance using a pair of ear-microphones [15]. Repeating this measurement for several excitation points we can collect a set of responses, each one containing all the information necessary to render a certain angular position for a sound source, and even its near-field distance from the subject.

These responses are called Head-Related Impulse Responses (HRIR). Their measurement has been standardized, along with the excitation point positions that must be chosen around the head [59, 60].

Although a careful, and necessarily subjective assessment of HRIRs results in a precise recording of all localization cues, both binaural and monaural, nevertheless a special effort has been done to understand how the HRIR structure depends on the subjective anthropometric parameters. Understanding those dependencies has led to *structural* models that, in principle, are able to synthesize HRIRs starting from certain subjective parameters, thus avoiding complicate individual recording sessions [4, 21].

On the other hand, analyses conducted on the spectral characteristics of the HRIR have helped in removing redundancies contained in such responses. This is helpful in the setup and management of a HRIR database, where the Fourier transforms of those responses, called Head-Related Transfer Functions (HRTF), can be efficiently encoded. This encoding is especially useful when the database must account for multiple distances per angular position [22]. Moreover, further simplifications in the HRTF content have been shown to be possible, by removing spectral components which are perceptually insignificant for auditory localization [87].

2.1.5 Objective attributes

In § 2.1.1 and § 2.1.2 we have seen spatial cues (both binaural and monaural) that are added to a sound as it moves from the listener's proximity to the ear canal

[2] in practice, a proper transfer function measurement technique must be chosen among those existing in the literature [57].

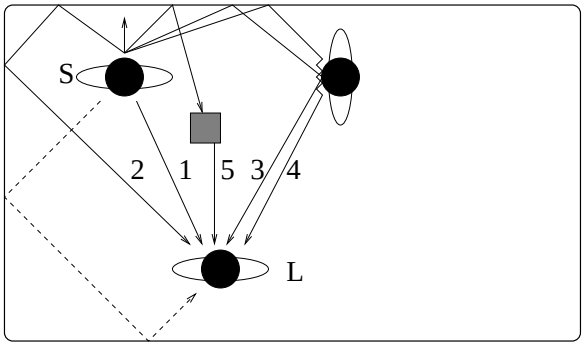

Fig. 2.1. Transmission of sound waves from a speaker (S) to a listener (L) in an echoic room.

entrance. *We are not going to model such cues.* Conversely, we are concerned with spatial cues coming from attributes that modify the sound during its journey from the source to the proximity of the listening point, regardless of the structural modifications to sound caused by the presence of the listener in the environment.

In other words, we will model only the spatial attributes that are already present in a sound as long as it is detected by a transparent listener (such as a small microphone) located somewhere in the scene. Clearly, distance cues such as those seen in § 2.1.3 come from attributes like those.

We will call such attributes *objective*, since they are not subjectively dependent. From a modeling viewpoint, this means that our models take care of sounds until they reach the subject's proximity. Any further processing stage will be object of investigation for researchers involved in the presentation of sounds to a listener. Chapter 6 will survey the problem of sound presentation with a special emphasis on headphone listening.

Returning in more detail to the example of the speakers inside the anechoic chamber, in that case objective cues depend on the speaker's distance from the listener, her orientation inside the chamber, and the positions and physical properties of other persons and objects standing inside the chamber.

Of course, objective cues become much more interesting if the auditory scene is provided by reflective surfaces. This is the case of a normal living room, a listening room, a concert hall and so on. A typical listening context, in which a speaker (S) indirectly talks to a group of people, is depicted in Figure 2.1. By looking at the numbers labeling each sound trajectory, in that context we can observe all the phenomena previously described in the anechoic case, together with wall reflections.

1. The listener (L) first receives *diffracted* acoustic waves from the speaker's head;
2. Then, he starts receiving waves partially *reflected* by walls...
3. ... and by other listeners;
4. He receives also *diffracted* waves (by another listener, in the given example);
5. Objects in the scene, such as the cube depicted in Figure 2.1, can *transmit* acoustic waves as well.

We will not study the perceptual effects of a sound wave diffracting over the listener's head (depicted in dashed line).

The more reflective the surfaces (i.e., the more *reverberant* the enclosure), the more elaborate and rich the spatial attributes conveyed to the listener.

2.2 Reverberation

Although from objective spatial attributes humans cannot extract cues accounting for precise quantitative information about the enclosure characteristics, nevertheless the amount of reverberation in the sound has clear perceptual effects, and can dramatically change the listening experience.

Simple applications of spatial audio include *stereo panning*, via the synthesis of amplitude differences between the left and right loudspeaker [15]; *transparent amplification*, exploiting the precedence effect mentioned in § 2.1.1 [130]; *distance rendering*, that reproduces the auditory perception of a sound source located at a certain distance from the listener, inside a room having simplified wall reflection properties [29].

Moore's *Room in a Room* model implements a geometrical model for the positioning of a virtual sound source inside a listening room provided with a set of loudspeakers. By means of that model, relative amplitudes and time delays are computed for each loudspeaker channel in order to recreate the virtual source [124].

Those applications ask for simple processing, requiring only loudness variations and/or fixed time delays for each reproduction channel. In this sense, all of them can be reproduced using general-purpose digital mixing architectures that can be easily found in any professional sound reproduction installation [78].

On the other hand they point out the existence of two alternative approaches to reverberation: the former, driven by the psychophysics of spatial hearing; the latter, reproducing the listening environment.

2.2.1 Perceptual effects of reverberation

The literature on psychoacoustics of reverberation specifies a multiplicity of adjectives to identify the characteristics of an auditory scene where the reflections of a pure sound are conveyed to a listener [10]. Most of the research done in the field deals with listening contexts where reverberation plays a key role in the final quality of the listening experience. This is the case of concert halls and recording studios. Less attention has been deserved to everyday-listening contexts.

Although researchers in general do not use a unique terminology for characterizing the various aspects of reverberation, a fact that has been taken for granted is that a reverberant environment is best characterized using *perceptual* attributes. So, for example, the size of a source is perceptually assessed in terms of *apparent source width*, and, likewise, the overall volume of a listening context translates in the perceptual parameter of *spaciousness* [67]. Many other perceptual parameters have been proposed and, probably, a unique dictionary describing the perceptual attributes of spatialized sounds will be never put together.

It is clear that there is no direct relationship between source size and apparent source width, or between room volume and spaciousness: both those perceptual attributes are the result of a series of characters pertaining to the sound emitted by the source, along with several spatial effects affecting that sound during its journey to the listener. For this reason it is not uncommon, for instance, to experience that some medium-sized rooms can evoke more spaciousness than larger rooms.

The origin of spatial cues, hence, must be found in the acoustic signal. Investigations have been conducted, aiming at finding relationships between the perceptual attributes of sounds and some structural properties of the signal that reaches the listener's ears. Such investigations led to several important results. In some cases, psychophysical scales mapping specific features of the audio signal into perceptual attributes have been found [58]. These investigations have an invaluable importance for the design of *artificial reverberators*.

2.2.2 Perceptually-based artificial reverberators

Artificial reverberators synthesize spatial attributes in such a way that, after artificial reverberation, anechoic (or *dry*) sources sound like if they were played in an echoic environment.

Although reverberation models take great advantage from the psychophysical studies on the perception of spatial attributes, top-quality artificial reverberators are considered such as pieces of art. One reason for this reputation is that they add some plus to the general rules that give the basics for the design a good reverberator [16]. Another reason is that, since perceptual attributes are subjective by definition, they can go beyond the hearing sensations experienced in real environments, creating new listening standards and, hence, redefining the state of the art. Finally, in almost all cases reverberators are designed to serve the world of music.

The perceptual design approach has an important advantage: perceptual attributes depend on the structure of the audio signals. Hence, reverberators which are designed using traditional signal processing structures (such as linear filters) can manipulate those attributes directly. So, for instance, the decay time of a sound signal, mainly responsible for spaciousness, is manipulated in a traditional reverberator tuning the decay time of some filters contained in it, regardless of the wall absorption properties that a listening room should have to convey sounds having the same decay time [35].

Moreover, this design approach leads to models that are computationally much lighter than reverberation models derived from the physics of real environments. This evidence has closed the door in the past to possible applications of models based on a physically-consistent (or *structural*) approach, which asks for modeling the environment characteristics instead of the perceptual attributes.

Going in more detail, perceptually-based digital reverberators reproduce the *early echoes* and the *late reflections* perceived by a listener located inside a reverberant environment. Those echoes are depicted in Figure 2.2 in the form of delayed repetitions of an initial impulse, as they appear in the typical representation of an impulse response taken from an artificial reverberator. The density, magnitude and distribution of the early echoes is mainly responsible for the timbral characteris-

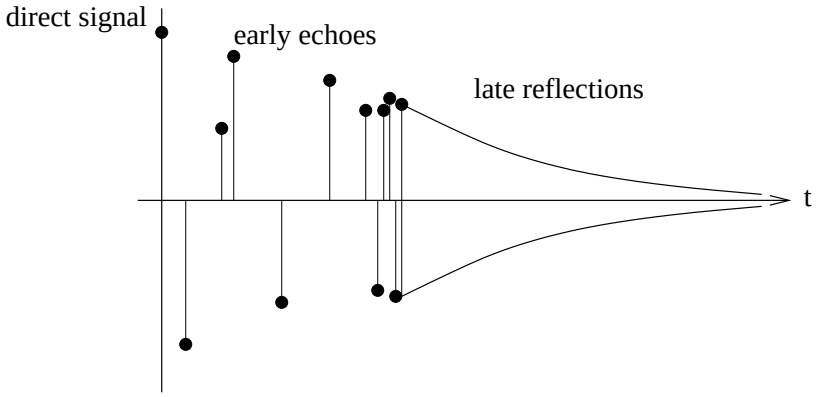

Fig. 2.2. Echoes appearing in the form of delayed repetitions of an initial impulse, in a representation of a typical impulse response of a digital reverberator.

tics of reverberation meanwhile the density and time length of the late reflections accounts for spaciousness, *warmth*, *envelopment* of the listening room [58].

The early echoes can be reproduced using generalized FIR filter structures, i.e., digital filters having transfer functions in the following form:

$$H(z) = \sum_{i=1}^{M} b_i z^{-k_i} \qquad , \tag{2.1}$$

H, in this case, reproducing M echoes of amplitude b_1, ..., b_M delayed by $k_1 T$, ..., $k_M T$ seconds after the direct signal, respectively (T is the sampling interval in the system).

On the other hand, late reflections are basically reproduced using filter networks capable of providing "dense" responses both in time and frequency [58,140].

Some interesting considerations can be also done for the user interfaces that can be found on board of artificial reverberators. Traditionally, those interfaces allow to choose the desired spatial effect through the selection of a mix of physical and perceptual attributes. For example, the user can choose among effects such as "large room" or "dry concert hall". Nevertheless, the future of these interfaces seems to be oriented toward a more intensive use of the perceptual parameters [78], further increasing the gap existing between the structural and perceptual design of artificial reverberators for musical purposes.

2.2.3 Physically-oriented approach to reverberation

The exponential nature of the physical problem that, starting from the description of a three-dimensional environment provided with acoustically reflective walls and objects, asks for finding out the set of echoes that reach a particular point in that environment after it has been excited by a sound source somewhere located, devises solutions whose computation explodes as soon as we try to calculate the positions and amplitudes of reflections that follow the early echoes. Models nevertheless exist that calculate that set of echoes with the desired accuracy.

Precise computations of the values taken in the spatial domain by the acoustic pressure function can be obtained using *boundary elements* and *finite elements* methods. These methods compute the three-dimensional pressure field in the discrete space and time [91], starting from its differential formulation along with boundary and initial conditions. Although their accuracy in solving the difference problem is considered far beyond the precision required in sound synthesis, nevertheless unsurpassed results have been obtained using finite elements in the audio-video simulation of auditory scenes describing dynamic ecological events involving the motion of solid objects [109].

The same problem is solved also by *finite difference* methods. Although less flexible in the formulation of the difference problem compared with boundary and finite elements, finite differences result in simplified numerical schemes that calculate the pressure function over a desired set of spatial locations [18].

So far, the methods which received particular attention by researchers in audio had to provide solutions suitable for the real time. In particular, *ray-tracing* [85] and *image* [5] methods are appreciated for their simplicity and scalability in modeling wave reflection. In fact, following alternative geometrical approaches, both of them allow to calculate the echoes that follow the direct signal: the number of reflections that affect an echo, during its journey from the source to the listening point, determines the complexity needed for the computation of that echo. Hence, both methods devise algorithms that provide solutions, whose accuracy in the description of the echoic part of the signal can be tuned proportionally to the available computational resources.

More recently, the importance of modeling diffracted acoustic waves has been demonstrated particularly in propagation domains where the sound source is occluded to the listening point. Studies on wave diffraction yielded efficient modeling strategies based on a geometrical approximation of the diffraction effect [155].

An alternative approach to modeling sound propagation resides in the so called *wave* methods. Those methods are based on the linear decomposition of a physical signal into a couple of *wave signals*, this being possible for several domain fields including pressure fields. Wave signals are then manipulated once the transmission and reflection properties of the propagation domain are locally known.

Those local properties are transferred into particular elements defined by such methods: *Wave Digital Elements* in the case of Wave Digital Networks [47, 112, 133], *Scattering Junctions* in the case of Digital Waveguide Networks (DWN) [147], *Transmission Line Matrices* (TLM) in the case of TLM methods [30]. Such elements receive, process and finally deliver wave signals in such a way that the global network of elements can be computed by a numerical scheme.

Different wave methods have many points in common. Fundamental relationships existing between Wave Digital Networks and DWN have been shown by Bilbao [13]. More in general, there are several basic relationships between wave and finite difference methods [13] and between finite differences and finite elements [116]. Although a unified vision of the numerical methods existing for the simulation of wave fields is far from being proposed, nevertheless more research should be done to figure out possible links existing among those methods, and to define a common framework where the performance and the computational complexity of each method can be compared.

In this work, we will design resonating and sound propagation environments using the *Waveguide Mesh*, a numerical scheme belonging to the family of DWN. Relationships with finite differences have been shown for many Waveguide Mesh realizations [13, 50, 158]. In Chapter 3 we will address several known properties of the Waveguide Mesh, along with some novel observations about the numerical behavior of those schemes that may turn to be useful during the simulation of wave propagation domains.

2.2.4 Structurally-based artificial reverberators

The design of a structural model, i.e., a model which is capable of synthesizing objective spatialization cues starting from a set of physical parameters of the auditory scene, has also been considered by researchers. In the past such a model had no chances to be implemented on a real-time application, and even today the resources that are generally needed to achieve the real time do not justify the realization of a structurally-based artificial reverberator.

If we derive (2.1) from the early response of a real listening environment [17, 121], then our application will follow an hybrid approach, i.e., structural for the early echoes, and perceptual for the late reflections [140]. The corresponding model, thus, defines a consistent relationship between the early echoes and the structure of a listening room [58].

In the case when the listening conditions vary in time, for instance in presence of changes in the position of the listening point, geometrical algorithms (such as ray tracing or image methods, introduced in § 2.2.3) can be used to compute runtime, without excessive complexity, the digital delays and amplitude coefficients contained in (2.1). This enables to reproduce some dynamic spatial cues, such as the ones evoked when the sound source gets occluded in consequence of changes occurring in the scene description [93].

In parallel with a physically-consistent simulation of the early echoes, efforts have been made to match the late reflections with the room characteristics. This should make the audio more veridical especially in complex VR scenes, where the source and listener positions can be changed dynamically: for instance, when the listener moves toward a more "open" location in the auditory scene. Since the real-time constraint suggests to adopt computationally light models, then special maps are needed which establish efficient relationships between the environmental parameters and the driving coefficients of the filter networks accounting for the late reflections.

Those maps must be layered in between the model and the interface. In this way the late reverberation is informed with a scene description. Clearly, such maps can be implemented once the relationships existing between late reflections and room characteristics are known. If those relationships derive from perceptual considerations, then the resulting reverberator once again represents a compromise between the perceptual and structural approach.

In any case such a "layered" approach leads to indirect communication between the reverberator and its interface. In spite of this, the most recent models for late reverberation (in particular those employing Feedback Delay Networks) translate into structures whose direct accessibility to certain control parameters envisions closer relationships with the physical world [79, 122].

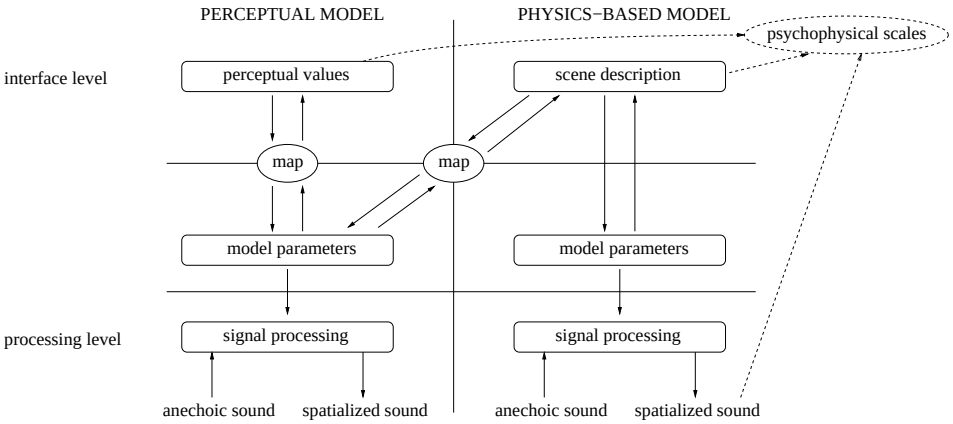

Fig. 2.3. Perceptual vs. physics-based approach.

2.3 Physics-based spatialization models

In this section we can outline in more detail the main scope of this work. We will call *physics-based spatialization model* a model capable of reproducing objective spatial cues (such as those defined in § 2.1.5) through the simulation of a listening environment whose physical/geometrical parameters are explicitly represented in the model.

Explicit parameter representation turns out to be a fundamental feature in our model. By means of this property we can keep those parameters simultaneously available at the interface level, as *control parameters*, and at the processing level, as *driving parameters*, with no intermediate stages connecting those two levels.

Numerical methods, such as those seen in § 2.2.3, are possible starting points for the development of physics-based spatialization models. In fact, by means of those methods we can design models in which:

1. signals are processed through elements, whose driving parameters are directly derived by the model;
2. the description of the listening environment is directly derived by the model as well.

In Chapter 3 we will see that Waveguide Meshes exhibit both properties, so that they are eligible for the design of physics-based models.

2.3.1 Physically-informed auditory scales

For what has been said in Section 2.3, a physics-based model looks attractive because of the direct representation of a listening environment existing inside it. Figure 2.3 gives a graphical explanation of the differences and relationships existing between the perceptual and physics-based approach.

The weak points, which cannot be addressed by the physics-based approach, reside in the aforementioned gap existing between physical acoustic measurements

and perceptual auditory scales, along with the computational requirements needed by physics-based models.

Concerned with the former point, *we propose to evaluate physics-based models perceptually, hence obtaining, using those models, psychophysical scales that are functions of physical spatial parameters*. This sounds quite obvious for certain scales: for example, loudness and pitch are functions of the sound pressure and frequency. On the other hand, the same functions become less obvious for other auditory spatial scales such as *presence*, *warmth*, *heaviness* and so on, for which only relationships with certain properties of the signal have been proposed instead, avoiding detailed physical interpretations of those scales [78, 166].

Although the definition of a physics-based psychoacoustics of spatial hearing is far beyond the scope of this work, nevertheless we will give examples in Chapters 4 and 5 showing that, due to their versatility and accuracy, physics-based models are becoming promising tools for the simulation of auditory scenes, at least for VR applications, and at most for psychophysical testing based on virtual listening environments.

Concerned with the latter point, *we hypothesize that humans' auditory perception of a listening environment translates into simplifications in the models*. Although we cannot prove that hypothesis, the experiment carried out in Chapter 5 suggests that physics-based models in practice are not required to simulate all aspects of the listening environment, but those which are significant for its auditory representation.

This idea of representing the auditory scene instead of the listening environment *in toto* envisages new horizons for physics-based spatialization, that remain still widely unexplored also by this work. Such environments will be captured by "spatial contexts" that, although including a reduced (however rearranged) set of physical and geometrical figures of the environment, nevertheless provide all the psychophysical dimensions needed to convey an exhaustive auditory scene description to a listener.

This kind of approach is already familiar to researchers working with physically-informed models for the synthesis of *ecological sounds* [62, 165]: those models use to capture only some macroscopic aspects of the physical process that generates a sound, whereas neglect or treat in some simple way more subtle factors acting during the process [34, 118, 125]. Experiments conducted using physically-informed models have demonstrated that, if correctly conceived, such models capture the most relevant perceptual aspects of the process, meanwhile they (probably) miss other aspects that don't produce strong auditory cues. Those models can be used to specify perceptual scales that are functions of the physical parameters, hence informing about "how do we hear in the world" [61].

Similarly, if successful a physics-based approach to spatialization would be able to explain auditory spatial impressions such as envelopment, warmth and so on by relating them to physical descriptions of the scene. Psychophysical auditory scales would come out from subjective judgments on sounds spatialized by physics-based models containing a precise description of the auditory scene.

This intuition is graphically highlighted by the dashed lines depicted in Figure 2.3. While envisioning that scenario, this work only investigates perceptual scales

having a clear and direct physical characterization: distance, that is the argument of Chapter 5, along with *roundness* and *squareness*, subjects of Chapter 4.

3

The Waveguide Mesh: Theoretical Aspects

"Per qualche motivo, la violazione della logica mi tocca sul piano morale."
[Carlo Ginzburg. Intervista televisiva sul libro
Il giudice e lo storico. Considerazioni in margine al processo Sofri. 2001.]

"For one reason or another, the violation of logic touches me on the moral side."
[Carlo Ginzburg. Television interview on the book
Il giudice e lo storico. Considerazioni in margine al processo Sofri. *2001.]*

As introduced in Section 2.3, the Waveguide Mesh [158] (from here called WM) can be used as a tool for the synthesis of physics-based models. This fact will be issued more precisely in this chapter, along with the theoretical limits that this approach to sound spatialization actually exhibits.

Ideal wave pressure fields are mathematically described by a differential equation defined over a *problem domain*. Solutions for this equation can be investigated after consistent initial and boundary conditions are set, in this way completing the formulation of the differential problem. For certain simple boundary conditions, closed-form solutions can be found [102].

Once the problem is addressed in that way, pressure fields take the form of a composition of waves that propagate along the problem domain. The wave decomposition of a multidimensional pressure function that solves the differential problem is a general result, in a way that we can study the evolution in time of a pressure function once we know the initial values assumed by its wave components and the physical laws that govern the evolution of those waves.

WMs, indeed, work in the way just described: by embedding elements that model wave scattering and wave reflection, they process signals that play the role of "wave" signals in the discrete time and space. In this way, WMs provide a numerical approximation of the pressure waves that propagate along a problem domain. As a direct consequence of this modeling approach, an approximation of the pressure function can be immediately figured out by the WM at any processing step.

An incorporation of the WM in the framework of DWN [147] has been recently made by Bilbao [13]. DWN allow to model several propagation fields, whose differential equations can be numerically computed using wave scattering methods.

Many differential problems that can be dealt with using DWN have practical application in acoustics and sound synthesis.

DWN have close connections with the so-called Finite Difference Time Domain methods, so that several properties holding for the latter family can be exported to the former one. In particular, WMs have a direct correspondence with Finite Difference Schemes [157]: this analogy enables to assess with precision, by means of a Von Neumann analysis [152], the error introduced by the WM in the approximation of the exact solution of the differential problem. The Von Neumann analysis can be extended to the various formulations [13,138] and geometries [24,50,160] of the WM, and informs about the accuracy provided by simulations that are made using the WM.

The reader interested in a deeper comprehension of the WM, in its several formulations, is encouraged to refer to Bilbao [13,14], where a rigorous formal definition of the WM in the framework of DWN is proposed, along with its numerical properties and relationships with Finite Difference Schemes.

Most of the analyses presented in the literature, evaluating the error introduced by a numerical scheme, do not deal with the spectral distortions that affect the output as a consequence of this error. As a rule of thumb, it is assumed in literature that numerical methods provide results whose accuracy decays with increasing frequency or, in other words, that such results are unreliable in the high frequency.

This is certainly true, due to the unavoidable finite precision of the model, along with the discrete (bandlimited) representation of the physical signals that are involved in the process. On the other hand, this lack of results sometimes puts the model designer in trouble when she must determine the granularity of the domain description, in order to obtain a certain accuracy in the simulations up to a specific frequency.

Another issue which is often neglected by numerical analysis is how to choose the temporal granularity of the simulations. That issue is usually a minor problem in most applications, since numerical schemes are typically worked out offline, in a way that oversampling them in time is not so critical in the majority of cases.

In the audio field, the real time is a major concern. This concern is even more compelling as long as spatialization is taken into account (refer to § 2.2.2). Hence, a precise knowledge of the consequences which follow by choosing a particular time step for a scheme is certainly informative for the designer of audio systems. First of all, it would give information about the spectral characteristics of the output signals. More in general, the timing effects should be under control when a real-time execution of a numerical simulation is planned on signal processors, or other types of architectures, that are capable of processing signals at a fixed sampling rate: in that case, the time step becomes a constraint in the problem, in a way that its effects must be clear to the designer since the beginning of the design activity.

We have conducted this kind of analysis for the WM. For sake of simplicity the analysis has been limited to WM geometries that model two-dimensional (2-D) problem domains, namely the *square*, the *triangular* and the *hexagonal* WM. Although, obviously, 2-D domains cannot account for three-dimensional (3-D) pressure fields, nevertheless the proposed analysis can be straightforwardly extended to the third dimension. Moreover, some conclusions that have been demonstrated

in the 2-D case hold also for the 3-D case: first of all, that WM grids containing gaps in between are proportionally less efficient than gap-free grids.

We complement the analysis presenting a curious interpretation of the square WM, that motivates from a physical viewpoint the peculiar symmetry exhibited by its frequency response, along with a method that, exploiting this symmetry, halves the number of computations that are required to calculate that response.

An important question concerning the designer of spatialization tools is how to control, in a WM, wave transmission and wave reflection for a particular problem domain. Once again, the waveguide approach yields a direct way to deal with those aspects of the problem. In fact, WMs allow a local selection of the *impedance*, by means of which the acoustic properties of materials can be completely specified. In this way the problem domain can be characterized. A thorough treatment on how to set impedances in the WM has been made by Bilbao [13].

Rather, in this work we focus on *boundary reflections*. In particular, we have studied how accurately they can be modeled using scattering. We have found evidence that boundary scattering models are more accurate, since their impedances account for the direction of the incident wave, and waves are reflected along directions that are consistent with the impedance value. Again, the analysis has been limited to 2-D propagation domains, although its conclusions may be qualitatively extended to the 3-D case, and quantitatively assessed via WM simulations of 3-D bounded propagation domains.

Finally, we propose a modified version of the WM that exhibits a high accuracy. A way for reducing the numerical error introduced by the WM has been investigated by several researchers involved in the subject, and finally a technique to manipulate this error has been proposed along with an offline method that performs that manipulation [137]. Nevertheless, a numerical scheme that realized that manipulation online had not been presented so far. Here we propose a method enabling online manipulation, then we construct a modified WM where the numerical error has been minimized.

This chapter embraces all the subjects just outlined. It is structured in the following way:

- a minimum set of basic concepts, necessary to understand the WM in its several formulations, is introduced;
- a spatial and temporal frequency analysis of the square, triangular and hexagonal WM is conducted. By means of that analysis the spectral properties of the responses coming out from those meshes are motivated;
- the previous analysis is complemented with further observations on the response of the square WM, and a technique for computing that response that halves the number of computations is proposed;
- a strategy to improve the accuracy in the simulation of the boundary is given;
- a numerical scheme which minimizes the numerical error is proposed, along with a technique to compute it.

3.1 Basic concepts

Here, we introduce some basic concepts of wave propagation in pressure and mechanic fields, together with their modeling using the WM.

3.1.1 Wave transmission and reflection in pressure fields and uniform plates

In a pressure field, the wave equation that expresses a deviation p from atmospheric pressure over a coordinate point (x, y, z) is written at time t as

$$\frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} = \frac{\partial^2 p}{\partial x^2} + \frac{\partial^2 p}{\partial y^2} + \frac{\partial^2 p}{\partial z^2} \quad . \tag{3.1}$$

This equation describes the case, considered representative enough for real situations, of *linear* pressure fields. The parameter c gives the speed at which pressure waves propagate along the field domain [102]. Linear pressure fields describe lossless, constant-impedance and constant-speed wave transmission domains.

A straightforward particularization of this equation to 2-D domains (that is, letting for example $\partial^2 p/\partial z^2 \equiv 0$) leads to a description of only a reduced set of 2-D wave propagation problems. In fact such problems (mainly referring to mechanical and electro-magnetic fields) are described by a more general partial linear differential equation system, that defines the *Parallel-Plate Transmission Line* problem. The parallel-plate transmission line problem, in particular, can account for local variations of the medium parameters (represented by *capacitance* and *inductance*), that induce corresponding variations in the impedance and loss characteristics of the problem domain, and in the wave propagation speed [13].

For our purposes, the case of constant capacitance and inductance is sufficient. As explained, this corresponds to analyzing equation

$$\frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} = \frac{\partial^2 p}{\partial x^2} + \frac{\partial^2 p}{\partial y^2} \quad , \tag{3.2}$$

that is clearly a reduced version of (3.1) to the 2-D case. This means that we will limit our 2-D case studies to problems which are described by (3.2). This restriction must not be considered as a limitation in this work, whose main scope is the modeling of 3-D linear pressure fields. In this perspective, the description of wave propagation problems via Equation (3.2) is particularly attractive, since it translates into a major simplification of the corresponding WMs compared to those modeling 3-D domains, without loss of generality of most of the conclusions coming out by modeling (3.2) instead of (3.1).

The presence of a boundary induces the phenomenon of wave reflection. We focus on the reflection of sound waves, giving some basic physical laws whose validity holds in the case of *locally reacting surfaces*, i.e., surfaces whose reaction to incident waves is pointwise, independently of the motion of neighboring surface points [90].

The impedance z is defined as the ratio between pressure and normal fluid velocity, $v^{(n)}$:

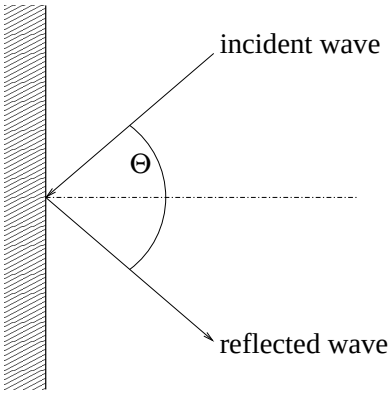

Fig. 3.1. Wave reflection.

$$z = \frac{p}{v^{(n)}} \quad . \tag{3.3}$$

In our case of interest, the surface reflects an incident pressure wave, p_i, in a way that each frequency component of the reflected wave, p_r, is both changed in amplitude and phase. Hence, reflection can be conveniently expressed in the frequency domain in terms of a complex ratio (called *reflection factor*) between the Fourier transforms of the reflected and incident wave, respectively P_r and P_i:

$$R(f) = \frac{P_r(f)}{P_i(f)} \quad . \tag{3.4}$$

In the case of real surfaces, reflection factors change to a broad extent, according to the material the surfaces are made of. Most of those surfaces exhibit a lowpass behavior with respect to wave reflection. In other words, they mainly absorb high frequency wave components.

It can be shown that relation

$$\frac{Z}{Z_0} = \frac{1}{\cos\Theta}\frac{1+R}{1-R} \tag{3.5}$$

holds between the (Fourier-transformed[1]) impedance Z of a surface point and its reflection factor. In (3.5), $Z_0 = \rho_0 c$ is the acoustic impedance of the transmissive medium (ρ_0 is the density of air), and Θ is the angle of incidence. Figure 3.1 explains wave reflection graphically in the case when the incident wave is reflected out from the surface along a single direction, forming with the normal an angle which is equal to the angle of incidence. This kind of reflection commonly takes place in optical mirrors.

3.1.2 WM models

Let us sample a P-dimensional domain into equal elements (for instance, volume elements in the case of a 3-D pressure field) having N facets, each one of those being

[1] in most of the literature the impedance is directly defined in the frequency domain.

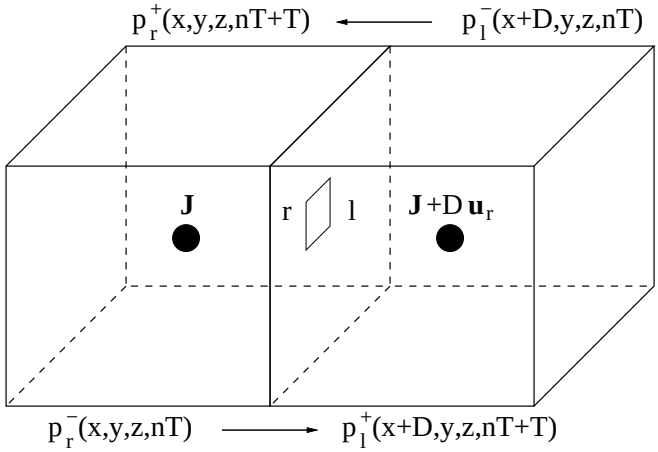

Fig. 3.2. Two adjacent six-facets elements forming a 3-D spatial domain. Those elements are centered around positions $\boldsymbol{J} = (x, y, z)$ and $\boldsymbol{J} + D\boldsymbol{u}_r = (x + D, y, z)$, respectively, and they have been modeled as scattering junctions. At time step nT output pressure waves, p_l^- and p_r^-, are scattered out from facets l and r: they respectively become incoming pressure waves, p_r^+ and p_l^+, for the opposite element at time step $nT + T$.

centered around a position $\boldsymbol{J} = (x_1, \ldots, x_P)$; let us model any of those elements as a *N-port scattering junction* [147] in a way that, at any time step, each element receives and scatters out one-dimensional waves through its N facets:

$$p_i^+(\boldsymbol{J}, nT)\,,\, p_i^-(\boldsymbol{J}, nT) \quad , \quad i = 1, \ldots, N \quad , \tag{3.6}$$

respectively. Figure 3.2 reports an example for a pressure field that has been spatially sampled into cubic elements: for each element, six pairs of input and output pressure waves move instantaneously in and out, respectively to and from the element. In that figure only two elements have been reported, along with their (two) mutual incoming/scattered waves.

Kirchoff's laws (namely, conservation of the instantaneous power) yield an instantaneous relation between the scattered and the incoming waves. The constancy of the medium characteristics among adjacent elements provides lossless transmission at a constant speed (i.e., *adaptation* between adjacent junctions), in a way that this relation simplifies to

$$p_i^-(\boldsymbol{J}, nT) = \frac{2}{N} \sum_{k=1}^{N} p_k^+(\boldsymbol{J}, nT) - p_i^+(\boldsymbol{J}, nT) \quad , \quad i = 1, \ldots, N \quad , \tag{3.7}$$

in which, without loss of generality, we consider pressure waves.

The WM model states that scattered waves take exactly one time step to reach an adjacent scattering junction. This makes the resulting scheme explicitly computable. In fact, referring once again to Figure 3.2, we have that two adjacent scattering elements facing each other through facets indexed r and l, respectively located at discrete-space coordinates $\boldsymbol{J}$ and $\boldsymbol{J} + D\boldsymbol{u}_r$, define the following relations:

$$p_r^+(\boldsymbol{J}, nT + T) = p_l^-(\boldsymbol{J} + D\boldsymbol{u}_r, nT) \tag{3.8a}$$

$$p_l^+(\boldsymbol{J} + D\boldsymbol{u}_r, nT + T) = p_r^-(\boldsymbol{J}, nT) \tag{3.8b}$$

where the *spatial step* D separates adjacent elements along the direction defined by $\boldsymbol{u}_r$, and the *temporal step* T separates subsequent discrete-time evolutions of the scheme.

From Equations (3.8a) and (3.8b) we observe that each scattering element obtains the values of its incoming waves from values previously scattered by adjacent elements. Those values are then used to compute new scattered waves at the current time step, via a simple computation such as (3.7). This scheme, hence, is inherently well-suited for implementation on parallel architectures. Although its implementation at audio sample rates on a current signal processing device, such as a DSP, limits the complexity of the scheme to few scattering junctions, on the other hand this complexity can be raised up to virtually no limits as soon as the scheme is embedded on some kind of modular architecture capable of providing, individually for each element, shifts as those described by Equations (3.8a) and (3.8b), and signal processing such as that seen in (3.7).

3.1.3 Digital Waveguide Filters

So far, we have seen a method that computes wave propagation along unbounded, ideal media as described by (3.1). We now outline the Digital Waveguide Filter (briefly, DWF), a block that models waveguide terminations and, for this reason, can be used to bound a WM [159]. Later in this chapter we will see how we use DWFs to model the boundary of a propagation domain.

Equation (3.4) suggests that, in a WM, the reflection factor can be modeled by a digital filter, whose transfer function calculated between incoming and outgoing wave signals resembles R up to a certain frequency (at most the Nyquist frequency). We will label this discrete-time transfer function with the same symbol, deliberately confounding the reflection factor with its realization, in the discrete-time, obtained using a DWF. More precisely, the z-transform of this filter can be put in the following form [98]:

$$R(z) = \frac{b_0 + b_1 z^{-1} + \ldots + b_M z^{-M}}{1 + a_1 z^{-1} + \ldots + a_P z^{-P}} \quad . \tag{3.9}$$

In the case when we are modeling, instead of a multi-dimensional pressure field, a network of waveguides each one having its own impedance, Equation (3.7) can be generalized to account for those individual impedances. In fact, considering N ideally joint waveguides having impedances respectively equal to $z_1, \ldots z_N$, then Kirchoff's laws yield a formula that puts scattered waves in relation with incoming waves in the transformed domain [147]:

$$P_i^- = \frac{2\sum_{k=1}^{N} P_k^+ / Z_k}{\sum_{k=1}^{N} 1/Z_k} - P_i^+ \quad , \quad i = 1, \ldots, N \quad . \tag{3.10}$$

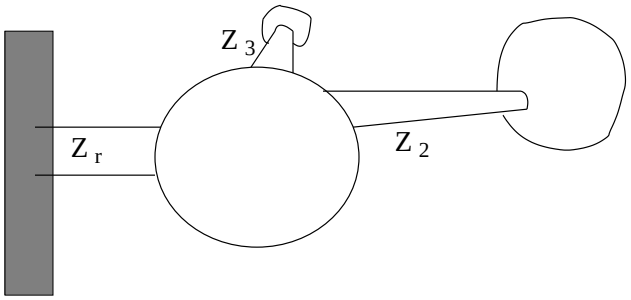

Fig. 3.3. Particular of a representation of a waveguide network. Three waveguides, each one having its own impedance, and three scattering junctions can be seen. One waveguide terminates against a reflective surface having impedance Z_r.

Figure 3.3 illustrates a particular in a representation of a waveguide network, where a junction leading to a reflective surface has been modeled to have an impedance equal to the surface impedance Z_r. Such a network can represent, for instance, a network of ideal pipes or a cable network. A WM is one particular realization of a waveguide network, made of digital waveguides having all the same impedance and unitary length.

Equation (3.10) allows to model junctions, in the WM, accounting for the boundaries in terms of their reflection factor. In fact (refer to Figure 3.3) it is sufficient to consider such "boundary junctions" as connecting one or more waveguides which are lumped in a medium having its own impedance, where waves cannot propagate. In that case Equation (3.5) can be substituted inside (3.10), in a way that the reflection factor takes part in the scattering process. To obtain this, we impose that boundaries neither accept nor reflect back waves to the junction (i.e., that signals both coming from and scattered to waveguides which are lumped in the boundary are null).

In the case of a 2-port scattering junction modeling the termination of an ideal one-dimensional air duct having impedance Z_0, (3.10) yields

$$P^- = \frac{Z_r - Z_0}{Z_r + Z_0} P^+ \quad , \tag{3.11}$$

where Z_r is the impedance at the termination. Hence, the reflection factor of the termination is equal to

$$R = \frac{P^-}{P^+} = \frac{Z_r - Z_0}{Z_r + Z_0} \quad . \tag{3.12}$$

If we explicit the impedances as function of the reflection factor in (3.12), we have

$$\frac{Z_r}{Z_0} = \frac{1+R}{1-R} \quad , \tag{3.13}$$

that equals (3.5) in this particular case, in which we have assumed $\Theta = 0$.

3.2 Spatial and temporal frequency analysis of WMs

The original question, leading to the analysis presented in this section, is the following one: let us simulate a WM model at a temporal step T. Where is the highest representative frequency component contained in the responses coming out from that simulation?

The *Nyquist frequency*, which is equal to half of the sampling frequency F_s:

$$\frac{F_s}{2} = \frac{1}{2T} \quad , \tag{3.14}$$

although being an absolute (and obvious) upper limit for the responses coming out from the model, does not give a general answer to that question. In fact, such responses have different upper frequency limits according to the mesh geometry. To demonstrate this, we must understand how the spatial traveling wave components are simulated by the model.

3.2.1 Temporal frequencies by spatial traveling wave components

The way of partitioning the propagation domain into scattering elements can be chosen between several *geometries* (for example, Figure 3.2 depicts a cubic geometry). As a constraint in this choice, we will limit to partitions where *all scattering elements have the same extension* (say, volume in the 3-D case). In practice, we will only deal with WMs where adjacent scattering junctions are separated by a common spatial step, D.

Each choice results in a different tessellation of the spatial domain. Figure 3.4 shows some possible geometries of the WM, both 2-D and 3-D.

Equation (3.7) shows that the pressure over a scattering element is completely described by the total incoming wave pressure. In fact, it can be seen that this sum gives the actual pressure value p_J associated to that element [158], that is, rearranging (3.7):

$$p_J(\boldsymbol{J}, nT) = p_i^+(\boldsymbol{J}, nT) + p_i^-(\boldsymbol{J}, nT) = \frac{2}{N}\sum_{k=1}^{N} p_k^+(\boldsymbol{J}, nT) \quad \text{for each } i \quad . \tag{3.15}$$

This means that the information on the pressure function which can be extracted by a WM at any time step is made of spatial sample values, each one accounting for the corresponding scattering element. Hence, borrowing concepts from the Multi-Dimensional Sampling Theory, we can measure that information using *multi-dimensional sampling lattices* [25, 42].

Sampling lattices describe the set of points where a discrete multi-dimensional signal is defined. More precisely, a sampling lattice in a P-dimensional domain (usually the continuous domain $\mathcal{R}^P$) is the set of points

$$L(D) = \{\boldsymbol{L}(D)[n_1 \ \ldots \ n_P]^T \, , \, (n_1, \ldots, n_p) \in \mathcal{Z}^P\} \quad , \tag{3.16}$$

where $(n_1, \ldots, n_p)$ is a P-tuple of signed integers, and $\boldsymbol{L}(D)$ is a non-singular $P \times P$ matrix containing a free parameter D. The symbol ^{T} denotes transposition.

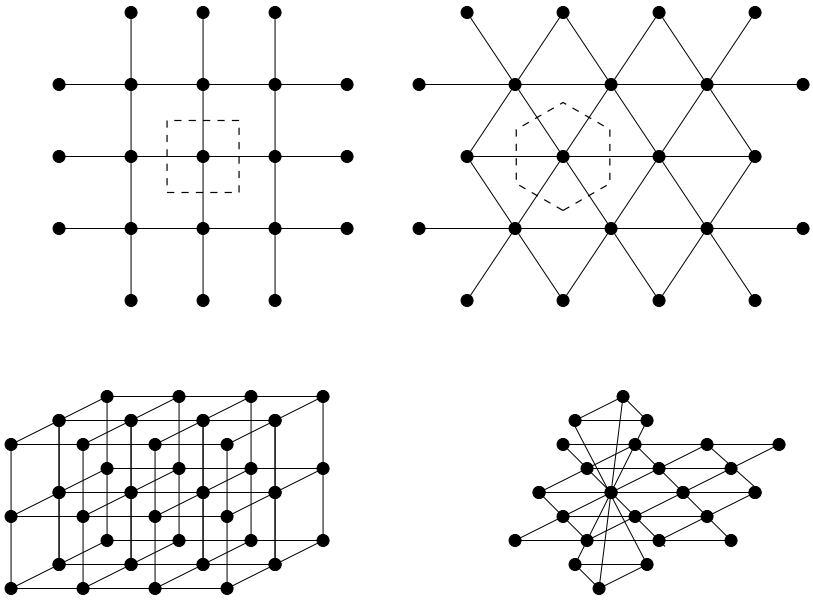

Fig. 3.4. WM geometries. Rectangular (above, left), triangular (above, right), cubic (below, left), 3-D triangular (below, right). 2-D geometries show also the corresponding scattering elements (enclosed in dashed line).

Sampling lattices include a notion of *density* $\mathcal{D}$ that accounts for the extension of the elements: the bigger the element, the lower the density is. Quantitatively:

$$\mathcal{D}_{L(D)} = \frac{1}{\det \boldsymbol{L}(D)} \quad . \tag{3.17}$$

Keeping in mind Figure 3.2 once more, in that example the scattering junctions lie on positions defined by the set

$$L_C(D) = \{D\boldsymbol{I}\,[n_x\, n_y\, n_z]^T\,,\, (n_x, n_y, n_z) \in \mathcal{Z}^3\} \quad , \tag{3.18}$$

where $\boldsymbol{I}$ is the 3×3 identity matrix and D is the spatial step chosen for that WM model. In particular, it is $\mathcal{D}_{L_C(D)} = 1/D^3$.

The Multi-Dimensional Sampling Theorem specifies the (P-dimensional) support, $L^*(D)$, containing the frequency components $\boldsymbol{\xi}$ of a signal defined over $L(D)$. In this way, we can precisely assess the spectral information provided by the discrete-space function defined by the WM at any time step. For example, in the case of the 3-D rectangular WM the spatial frequencies of the discrete pressure function $p_J(\boldsymbol{J}, nT)\,,\, \boldsymbol{J} \in L_C(D)$, observe at any time step the following condition on the support:

$$\boldsymbol{\xi} = |\xi_x \;\; \xi_y \;\; \xi_z|^T \,:\, |\xi_x| < \frac{1}{2D}\,,\, |\xi_y| < \frac{1}{2D}\,,\, |\xi_z| < \frac{1}{2D} \quad . \tag{3.19}$$

such that $L_C^*(D)$ is a cube of volume $1/D^3$ centered around the spatial frequencies origin.

The speed of propagation of a numerical wave component traveling along a WM is a function of its own spatial frequency. We express this calling $c(\boldsymbol{\xi})$ this propagation speed, when traveling wave components have spatial frequency equal to $\boldsymbol{\xi}$. The dependency of the propagation speed on spatial frequency is part of the numerical error introduced by the model: in fact, contrarily to that, Equation (3.1) envisages a common speed for any component.

This error is known in literature as *dispersion* and, as mentioned in the beginning of this chapter, can be precisely calculated by conducting a Von Neumann analysis on the scheme [152]. As a direct consequence, the Von Neumann analysis highlights about the value of $c(\boldsymbol{\xi})$ for any WM geometry. On the other side, by conducting a stability analysis, condition

$$c(\boldsymbol{\xi}) \leq \frac{1}{\sqrt{P}} D F_s \tag{3.20}$$

is found to hold for the propagation speed in a WM modeling a P-dimensional propagation domain [152].

Once $c(\boldsymbol{\xi})$ is known over $L^*(D)$, we can finally answer the initial question. In fact, spatial frequencies $\boldsymbol{\xi}$ are turned into corresponding temporal frequencies f by multiplying the propagation speed times the spatial frequency magnitude:

$$f = c(\boldsymbol{\xi}) \sqrt{\boldsymbol{\xi}^T \boldsymbol{\xi}} \quad , \tag{3.21}$$

so that the upmost frequency contained in the response of a WM is equal to

$$f^{(\max)} = \max_{\boldsymbol{\xi} \in L^*(D)} c(\boldsymbol{\xi}) \sqrt{\boldsymbol{\xi}^T \boldsymbol{\xi}} \quad , \tag{3.22}$$

where the maximum is calculated over the support where the spatial frequencies are defined.

Following with the example of the 3-D rectangular WM, in that case the Von Neumann analysis states that the shortest spatial wave components also travel at the highest propagation speed. Thus, by (3.22) we obtain immediately

$$f_C^{(\max)} = \frac{1}{\sqrt{3}} D F_s \sqrt{\left(\frac{1}{2D}\right)^2 + \left(\frac{1}{2D}\right)^2 + \left(\frac{1}{2D}\right)^2} = \frac{F_s}{2} \quad . \tag{3.23}$$

Hence, that particular WM geometry produces responses whose spectral information matches the Nyquist value.

Starting from this overview on the frequency analysis of the WM, and its exemplification to the 3-D rectangular geometry, the reader is referred to Paper C for a comprehensive treatment of the 2-D triangular topology. In that work the hexagonal geometry is issued as well, as an example of a domain partition where the scattering elements have the same volume but different orientations.

From that treatment, two main conclusions are reported here:

1. compared with the rectangular geometry, the triangular WM processes more efficiently signals whose spatial frequency support is symmetrical around the origin of the frequency axes. Such signals are created, for example, by excitations having omni-directional directivity;

2. WMs containing gaps in their grids of scattering junctions (such as, for example, the hexagonal WM) produce responses that are proportionally less informative compared to those coming from gap-free mesh grids.

These two considerations can be qualitatively generalized to both 2-D and 3-D WMs. Quantitative figures can be calculated for the geometries that have not been explicitly considered in this work, by repeating the spatial frequency analysis starting from the related sampling lattice, and from the results provided by the corresponding Von Neumann analysis.

3.2.2 The frequency response of rectangular geometries

Equation (3.23) proves that 3-D rectangular WMs respond up to the Nyquist frequency. Repeating the calculation for the 2-D case is even easier. Nevertheless, rectangular geometries are characterized by responses that mirror at half of the Nyquist frequency.

This characteristic can be motivated reconducting the z-transform of those responses to a peculiar numerical property of the scheme [158]. Equivalently, it can be explained in the untransformed domain provided that the discrete-time pressure function defined on $L(D)$ observes the following two properties:

$$p_J(\boldsymbol{J}, 2nT) = 0 \quad , \quad \boldsymbol{J} \in L_1(D) \tag{3.24a}$$

$$p_J(\boldsymbol{J}, 2nT + T) = 0 \quad , \quad \boldsymbol{J} \in L(D) \setminus L_1(D) \tag{3.24b}$$

in which the sampling lattice $L_1(D)$ is a subset of $L(D)$, having density $\mathcal{D}_{L_1(D)}$ equal to half the density of $L(D)$. Properties (3.24a) and (3.24b) imply that any response, which is taken in correspondence of a scattering junction, presents sample values that are interleaved, in time, with null values. This results in mirroring at half Nyquist in the temporal frequency domain.

Rectangular geometries exhibit this characteristic as soon as an initial excitation is formed by setting the scattering junctions with initial pressure values, $p_J(\boldsymbol{J}, 0)$, that respect property (3.24a): in this way, any scattering element in $L_1(D)$ delivers null wave signals. At time step T those signals reach the adjacent elements, which belong to $L(D) \setminus L_1(D)$. Those elements have just delivered (in general) non-null signals at time step 0 and are receiving null signals at time step T, hence their pressure values become null in their turn. In this way at time $t = T$ the roles of the junctions belonging to $L_1(D)$ and $L(D) \setminus L_1(D)$ are exchanged. Clearly, this kind of swapping goes on indefinitely, in a way that a signal coming out from any point of the WM contains non-null samples interleaved with null values. Thus, properties (3.24a) and (3.24b) are satisfied.

Since the response of a WM is obtained initializing the system with a function, the *impulse* [98], that respects (3.24a) at $t = 0$, then that response will mirror at half Nyquist wherever we pick it up from the WM.

It is interesting to look for a modified version of the rectangular scheme that produces responses whose spectral content is limited to half of the Nyquist frequency. In fact we expect that such a modification, avoiding redundant computations, leads to a more economic way of computing those responses compared with the original rectangular WM. In practice we must look for a modified WM scheme

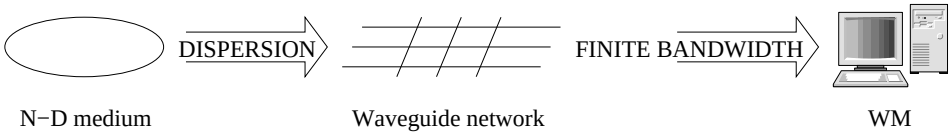

Fig. 3.5. Simulation errors produced by the reformulation of multi-dimensional propagation domains into waveguide networks and digital WMs.

that avoids processing the spatially interleaved null samples, meanwhile dealing with non-null samples in the same way the original WM does.

The reader is sent to Paper D, where a modified scheme working in the way just outlined is presented, together with figures that confirm that it requires less (more precisely, half) operations and memory occupation to compute a response. Similar results have been independently obtained also by Bilbao, who emphasizes different aspects in those schemes [12, 13].

The sampling lattice described by the modified rectangular WM resembles some sampling schemes used for the discretization of motion pictures. Those schemes efficiently encode the multi-dimensional information by interleaving the image samples both in time and space [25].

3.2.3 WMs as exact models of cable networks

We conclude the spatial and temporal frequency analysis of the WM giving an interesting characterization, that confirms that dispersion is introduced as soon as we simulate propagation along infinite directions using a model that can account only for a finite number of directions.

In § 3.1.3 we have mentioned that WMs are, in fact, numerical models of special networks of waveguides, in which all waveguides have the same length and impedance. Those networks model, for instance, compositions of air ducts or ideal strings in which each waveguide element is ideally joint with other waveguide elements. Also, we have seen that WMs can model multi-dimensional homogeneous propagation domains, to an extent that is quantitatively assessed by the Von Neumann analysis in terms of dispersion error, and by the frequency analysis in terms of upper frequency limit of the simulations.

A WM reduces to the Digital Waveguide [149] as soon as we particularize it to the 1-D case. In that case dispersion disappears, suggesting that that kind of numerical error comes into play only if we want to model multi-dimensional wave propagation. Then, the question is: are WMs responsible of dispersion, or, conversely, is dispersion the general consequence of a waveguide approach to multi-dimensional wave propagation?

The same question has been reformulated in Figure 3.5, where we have also anticipated a possible answer: while dispersion is introduced by the reformulation of the multi-dimensional propagation domain into a waveguide network, on the other hand a WM modeling that network causes the number of traveling wave components to become finite.

If that answer is true, then we can reconduct the differences existing between the response coming from an ideal propagation medium, and the response taken

from the WM, to two independent causes: dispersion artifacts, caused by discretization in the number of directions of propagation, and finite bandwidth, caused by discretization in space and time.

We will be able to demonstrate this only in the particular case of 2-D rectangular waveguide meshes simulating rectangular, ideally bounded domains, such as rectangular membranes ideally clamped at the edges. In order to do this we introduce *orthogonal cable networks*, that are orthogonal sets of interconnected cables modeled as ideal strings, simply *supported* (i.e., ideally clamped) at the ends. Those networks have been first studied in the field of applied mechanics. Refer to the 2-D rectangular WM of Figure 3.4 for a graphic description of the orthogonal cable network.

Before an exact solution of the cable network problem had been found, curiously the preferred modeling approach to that problem was the *membrane analogy* [146]: by means of it, the (analytical or simulated) solution provided by a physical membrane resembling the cable network according to certain specifications was used to approximate the cable network problem solution. That approximation had been found to be as more accurate, as larger the number of cables was; in any case, the membrane analogy provided an accurate solution as long as only the first (low frequency) vibration modes were considered.

Compared to the WM approach, the membrane analogy establishes a "backward link" between homogeneous, omni-directional propagation media and their modeling using waveguide networks. In fact, the solution provided by an omni-directional medium is used to approximate a network where waves can propagate along a finite number of directions. Clearly, the membrane analogy results to be particularly accurate in the simulation of the first vibration modes [146], exactly as WMs well approximate an omni-directional medium in the low frequency [13, 50].

The first exact solution of the cable network problem is dated 1985 [37]. That solution makes use of the *transfer matrix method*: it comes out by first calculating a transfer matrix for each single section of cable, then composing those matrices to successively obtain the transfer matrix for a joint, a chain, and finally the entire network.

The equations governing the displacement w in an orthogonal cable network are synthetically proposed here, in the case of cable networks having the same *pretension*, *length* and *mass per unit length* along both (x and y) directions. Such equations consist of [27]:

1. one *natural vibration equation* for each individual (both x- and y-oriented) section (i, j) of cable connecting two joints,

$$\frac{\partial^2 w_{(i,j)}}{\partial x^2} + \frac{\omega^2}{c^2} w_{(i,j)} = 0 \quad \text{or} \quad \frac{\partial^2 w_{(i,j)}}{\partial y^2} + \frac{\omega^2}{c^2} w_{(i,j)} = 0 \quad , \tag{3.25}$$

   where $\omega = 2\pi f$ is the *pulsation* of a vibration at frequency f, and c is the speed at which waves propagate along a cable section;
2. a condition of *continuity* across each joint, located on position $\boldsymbol{J}$,

$$w_{(i_1,j_1)}(\boldsymbol{J}) = w_{(i_2,j_2)}(\boldsymbol{J}) = w_{(i_3,j_3)}(\boldsymbol{J}) = w_{(i_4,j_4)}(\boldsymbol{J}) \quad \text{for all joints} \quad , \tag{3.26}$$

   where $(i_1, j_1), \ldots, (i_4, j_4)$ are cable sections connected together in a joint, including particular joints located at the cable ends;

3. a condition of *equilibrium* for all *internal nodes*,

$$\frac{\partial w_{(i_2,j_2)}}{\partial x} - \frac{\partial w_{(i_1,j_1)}}{\partial x} + \frac{\partial w_{(i_4,j_4)}}{\partial y} - \frac{\partial w_{(i_3,j_3)}}{\partial y} = 0 \quad , \tag{3.27}$$

   where $(i_1, j_1), \ldots, (i_4, j_4)$ are cable sections joint together in an internal node, such that (i_2, j_2) follows (i_1, j_1) along the x-direction and (i_4, j_4) follows (i_3, j_3) along the y-direction;
4. conditions of simple support at all cable ends.

This problem has been solved in closed form by using a transformation technique known as the *double U-transformation*. This technique is powerful enough to yield analytical solutions also in the case of beam networks [27,44] and orthogonal cable networks subjected to moving forces [28] or having periodically distributed supports [26].

In our case of interest, the pulsations of the modes vibrating in a square cable network made of $N-1$ cables, having length L, are the solutions of the equation [27]

$$\cos\frac{\pi\omega}{N\omega_0} = \frac{1}{2}\left(\cos\frac{\pi r}{N} + \cos\frac{\pi s}{N}\right) \quad , \quad r, s = 1, \ldots, N \quad , \tag{3.28}$$

as N is the number of sections forming each cable. In that equation it is

$$\omega_0 = \frac{\pi c}{L} \quad . \tag{3.29}$$

A cable network is a waveguide network. In fact, Equations (3.25) hold for a waveguide where waves propagate at speed c along it. Moreover, Equations (3.26) and (3.27) hold in particular in a series connection of waveguides [13]. Hence, a waveguide network provided with boundary conditions that simulate a rigid termination equals a simply supported cable network. From that it comes out that a square waveguide network sized $L \times L$, whose waveguides have a length equal to D in a way that it is $L = ND$, must vibrate at the pulsations given by Equation (3.28), which can be rewritten as

$$\frac{\pi\omega}{N\omega_0} = \arccos\left[\frac{1}{2}\left(\cos\frac{\pi r}{N} + \cos\frac{\pi s}{N}\right)\right] \pm 2k\pi, \quad r, s = 1, \ldots, N, \quad k \in \mathbb{N} \quad . \tag{3.30}$$

We now move to the discrete time and space, i.e, we change the waveguide network into a rectangular WM. This corresponds to substituting the waveguides in the network with unit-length Digital Waveguides. Those new elements cannot vibrate but at one frequency, so that their pulsation is unique, i.e., $k = 0$. Moreover they force waves to travel at speed $c = D/T$, where T is, as usual, the temporal step of the simulation.

Then, by (3.29) we have

$$\frac{\pi\omega}{N\omega_0} = \frac{\pi\omega}{N}\frac{ND}{\pi c} = \frac{\omega D}{c} = \omega T = 2\pi f T = 2\pi\frac{f}{F_s} \quad , \tag{3.31}$$

so that we finally obtain the modal frequencies in the WM:

r	s	f	r	s	f	r	s	f	r	s	f
1	1	0.0625	4	2	0.1925	4	4	0.2500	4	6	0.3075
1	2	0.0982	1	5	0.2064	5	3	0.2500	6	4	0.3075
2	1	0.0982	5	1	0.2064	6	2	0.2500	5	5	0.3125
2	2	0.1250	3	4	0.2194	7	1	0.2500	4	7	0.3264
1	3	0.1367	4	3	0.2194	2	7	0.2673	7	4	0.3264
3	1	0.1367	2	5	0.2241	7	2	0.2673	5	6	0.3417
2	3	0.1583	5	2	0.2241	3	6	0.2759	6	5	0.3417
3	2	0.1583	1	6	0.2327	6	3	0.2759	5	7	0.3633
1	4	0.1736	6	1	0.2327	4	5	0.2806	7	5	0.3633
4	1	0.1736	1	7	0.2500	5	4	0.2806	6	6	0.3750
3	3	0.1875	2	6	0.2500	3	7	0.2936	6	7	0.4018
2	4	0.1925	3	5	0.2500	7	3	0.2936	7	6	0.4018
									7	7	0.4375

Table 3.1. Positions in frequency of the vibrating modes of a rectangular WM sized 8×8 with rigid boundary terminations.

$$f = \frac{F_s}{2\pi} \arccos\left[\frac{1}{2}\left(\cos\frac{\pi r}{N} + \cos\frac{\pi s}{N}\right)\right] \quad , \quad r, s = 1, \ldots, N-1 \quad , \tag{3.32}$$

where the cases $r = N$ and $s = N$ have been excluded, since Digital Waveguides made of N pairs of unit delay elements, rigidly clamped at both ends, cannot vibrate but at $N - 1$ frequencies. By the way, the case $r = s = N$ results in frequencies that are equal to the Nyquist frequency (3.14).

As an application example let us consider a rectangular WM sized 8×8, sampled using a unitary temporal step. Equation (3.32) yields 49 modal frequencies, that have been reported in Table 3.1 in order of increasing frequency, together with the corresponding values r and s. Those values are also compared, in Figure 3.6, with the modal frequencies taken from a simulation of the same WM that has been excited in correspondence of a corner, and whose response has been acquired in correspondence of the opposite corner.

It can be seen that the correspondence between the calculated positions in frequency of the vibration modes, and the positions of the same modes when they are generated by a simulation, is excellent. This result is consequence of the analogy existing between rectangular WMs and discretized orthogonal cable networks.

It should not be difficult to extend the same analogy to the third dimension. Rather, an extension to non-orthogonal cable networks, aiming at finding analogies for example with 2-D triangular WM geometries, seems to ask for a good practice with the double U-transform.

3.3 Scattered approach to boundary modeling

As we have introduced in § 3.1.3, the simulation of a reflective and/or absorbing boundary turns out to be quite straightforward if DWFs are used to terminate the branches in a WM, in correspondence of the domain boundary. In particular, we have seen that boundary conditions, expressed as a variation in the impedance of

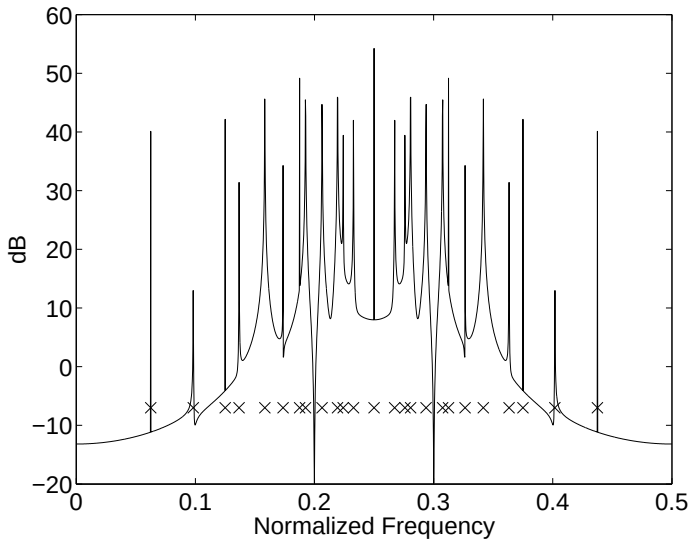

Fig. 3.6. Frequency response from a rectangular WM sized 8×8. ($\times$): frequency values obtained for the same WM using (3.32).

the medium along the boundary, can be embedded inside special 2-port scattering nodes, provided that the wave signals coming back and forth the reflective medium are permanently set to zero in those nodes.

In this section we deal with the issue of boundary modeling in further detail, starting from the following question: how can we model multiple reflections simultaneously occurring at the same point? In other words, what if the boundary junctions include more than one port facing the domain of propagation? This happens for example in the triangular geometries, both 2-D and 3-D (see Figure 3.4).

The analysis presented in § 3.1.3 is limited to single reflections, i.e., to 2-port scattering junctions. As a starting point, that analysis is sufficient to treat multiple reflections individually in a way that they can be modeled one by one, using 2-port scattering junctions (equivalent to DWFs) independently working one from the other, hence avoiding any kind of communication between wave signals coming from different directions. Figure 3.7 (left) illustrates this situation in the case of two incoming wave signals.

Alternatively, we note that Equation (3.10) accounts in general for N ports. Hence, by that equation we can immediately extend boundary reflection to $N-1$ incoming wave signals (Figure 3.7, on the right side, shows this for $N = 3$).

We have not proved that the "scattered" approach to boundary reflection leads in general to more accurate simulations. In spite of that, its application to cases where we can compare the model with the behavior of a physical boundary shows that the scattered approach to boundary modeling, in those cases, models wave reflection more naturally. As a consequence of that, simulations adopting scattering junctions at the boundary provide more accurate positioning of the modal frequencies.

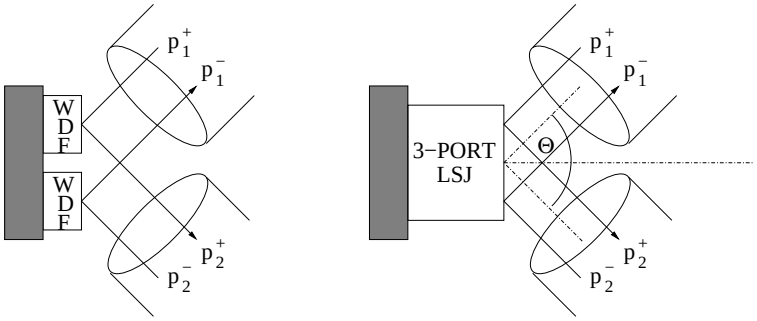

Fig. 3.7. Individual processing of two incident wave signals by means of DWF transfer functions (left). Integrated processing of two incident wave signals by means of boundary scattering junctions (right).

In the following of this section we will outline a scattered reflection model. During the exposition we will also review the DWF method for the design of boundaries in the WM [159] having a consistent physical interpretation [90]. The organization of this section took advantage also from the work of Fabio Deboni [36].

3.3.1 DWF model of physically-consistent reflective surfaces

The reflection factor R can be measured for many materials, as a function of the frequency of the incident wave component [90]. Table 3.2 shows *absorption*

Material	125 Hz	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz
Hard surface (bricks, plaster...)	0.02	0.02	0.03	0.03	0.04	0.05
Slightly vibrating wall	0.10	0.07	0.05	0.04	0.04	0.05
Strongly vibrating wall	0.40	0.20	0.12	0.07	0.05	0.05
Carpet	0.02	0.03	0.05	0.10	0.30	0.50
Plush curtain	0.15	0.45	0.90	0.92	0.95	0.50
Polyurethane foam	0.08	0.22	0.55	0.70	0.85	0.75
Acoustic plaster	0.08	0.15	0.30	0.50	0.60	0.70

Table 3.2. Absorption coefficients $\alpha = 1 - R$ at different frequencies for some materials (from Kuttruff [90]).

coefficients, defined as the complement to unity of the corresponding reflection factors, for several materials at different frequencies. From that table it can be seen that most of the surfaces which are normally found in everyday listening environments absorb higher frequencies more rapidly, i.e., they exhibit a lowpass behavior.

That spectral characteristic can be reproduced with sufficient approximation by tuning the parameters of the 1-st order transfer function provided by a simple spring/damper system, such as the one shown in Figure 3.8. This system in fact reacts to an external pressure acting over its surface, with a force proportional to the spring compression x and the piston velocity $v = \dot{x}$ of the damper.

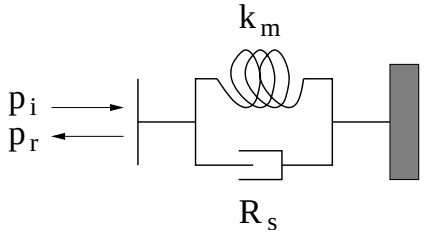

Fig. 3.8. 1-st order spring/damper system.

Then, considering a nominal unitary reflective surface area, it is $p(t) = k_m x(t) + R_s \dot{x}(t)$, where k_m and R_s are the spring constant and damping coefficient, respectively. After deriving with respect to the time variable, and Laplace-transforming p and v [80], the pressure/velocity transfer characteristics of the system is found out in the Laplace variable s:

$$\frac{P(s)}{V(s)} = \frac{k_m}{s} + R_s \quad . \tag{3.33}$$

Since, from (3.15), it is $p = p_r + p_i$ and $v = v^+ + v^-$, then using (3.3) we can substitute pressure and velocity values with wave pressure values, hence obtaining a transfer characteristic between pressure waves respectively coming out from and going to the boundary [159]:

$$\frac{P_r(s)}{P_i(s)} = R(s) = \frac{k_m/s + R_s - Z_0}{k_m/s + R_s + Z_0} \quad , \tag{3.34}$$

where $R(s)$ is the Laplace-transformed reflection factor of the spring/damper system. Recalling (3.12), from (3.34) it clearly descends that the boundary impedance modeled by our system is equal to

$$Z_r = k_m/s + R_s \quad . \tag{3.35}$$

A commonly used map, which transforms a transfer characteristic expressed in the Laplace domain into a discrete-time transfer function, is the *bilinear transformation*. This map preserves order and stability, and approximates the original transfer characteristic with an error that approaches zero as long as the frequency decreases [110]. The bilinear transformation requires to rewrite the Laplace variable according to the following map:

$$s \longleftarrow h\frac{1 - z^{-1}}{1 + z^{-1}} \quad , \tag{3.36}$$

where h is usually set to $2F_s$. By means of (3.36), we can finally formulate the transfer function of the DWF modeling the system shown in Figure 3.8:

$$\frac{P_r(z)}{P_i(z)} = R(z) = \frac{\frac{k_m - h(Z_0 - R_s)}{k_m + h(Z_0 + R_s)} + \frac{k_m + h(Z_0 - R_s)}{k_m + h(Z_0 + R_s)} z^{-1}}{1 + \frac{k_m - h(Z_0 + R_s)}{k_m + h(Z_0 + R_s)} z^{-1}} \quad . \tag{3.37}$$

This formula gives the coefficients that a 1-st order DWF must have to reproduce the damper/spring system, with a simulation accuracy specified by (3.36).

A DWF, whose transfer function equals (3.37), can be tuned in its coefficients to model some materials as those seen in Table 3.2. In particular, Table 3.3 shows

Material	k_m	R_s	Z_0
Hard surface (bricks, plaster...)	30.25	69.40	414
Carpet	9.42	6.49	414
Acoustic plaster	0.56	3.78	414

Table 3.3. Values assumed by the physical parameters k_m, R_s and Z_0 for the simulation of some materials ($F_s = 8$ kHz).

values that are assigned to the physical parameters k_m and R_s in such a way that the DWF responses model the characteristics of some of the materials appearing in Table 3.2, when the sampling frequency is set to 8 kHz and the air density is kept constant.

At this point we can terminate the WM boundaries with DWFs, whose coefficients are specified according to the absorption properties of the modeled surface. In other words, our DWF terminations provide a 1-st order model of a locally reacting surface, whose properties can be locally set by selecting the filter coefficients according to (3.37).

3.3.2 Scattering formulation of the DWF model

For what we have seen at § 3.1.3, the scattering formulation of a Digital Waveguide boundary termination translates into an input/output function, (3.12), that equals the DWF transfer function (3.34). In the meantime, (3.10) establishes a more general relation involving, in principle, the presence of $N-1$ incident waves to a reflective point plus the interface to the boundary. Since we work in the hypothesis of homogeneous propagation domains, the impedances appearing in (3.10) will be equal to Z_0 for all the components, except for the N-th one, which will account for the boundary impedance Z_r.

Hence, given the coincidence of the approaches in the singular and orthogonal reflection case, the question now is whether or not the scattered approach results in a more general modeling of the phenomenon of wave reflection.

We rewrite Kirchoff's laws in the case when incoming waves incide onto a reflective surface with angles $\Theta_1, \ldots, \Theta_{N-1}$ (refer to Figure 3.1). In that case, considering (as usual) pressure fields, the sum of the air velocities normal to the surface is null:

$$\sum_{k=1}^{N} v_k^{(n)} = \sum_{k=1}^{N} v_k \cos\Theta_k = 0 \quad . \tag{3.38}$$

Assuming this, then Equation (3.10) generalizes into the following formula [36]:

$$P_i^- = \frac{2\sum_{k=1}^{N} \frac{P_k^+ \cos\Theta_k}{Z_k}}{\sum_{k=1}^{N} \frac{\cos\Theta_k}{Z_k}} - P_i^+ \quad , \quad i = 1, \ldots, N \quad . \tag{3.39}$$

Note that if it is $N = 2$, then (3.39) equals Equation (3.5). Though, this case is not useful in practice, since the reflected wave is constrained by the scattering junction to have a direction which is opposite to the direction of the incident wave.

The case $N = 3$ turns out to be particularly interesting, since it accounts for mirror reflections such as the one depicted in Figure 3.1. In that case, Equation (3.39) yields—recall also Figure 3.7 (right):

$$\begin{cases} P_1^- = \dfrac{2Z_r cos\Theta}{2Z_r cos\Theta + Z_0} P_2^+ - \dfrac{Z_0}{2Z_r cos\Theta + Z_0} P_1^+ \\ P_2^- = \dfrac{2Z_r cos\Theta}{2Z_r cos\Theta + Z_0} P_1^+ - \dfrac{Z_0}{2Z_r cos\Theta + Z_0} P_2^+ \end{cases} . \tag{3.40}$$

This equation particularizes, in the case of the ideal *rigid wall* ($Z_r = \infty$ or, equivalently, $R = 1$), to

$$\begin{cases} p_1^- = p_2^+ \\ p_2^- = p_1^+ \end{cases} , \tag{3.41}$$

and conversely, in the case of the ideal *soft wall* ($Z_r = 0$ or, equivalently, $R = -1$), to

$$\begin{cases} p_1^- = -p_1^+ \\ p_2^- = -p_2^+ \end{cases} . \tag{3.42}$$

Relations (3.41) and (3.42) have consistency with the corresponding physical phenomena. With N increasing, (3.42) extends to $p_k^- = -p_k^+$, $k = 1, \ldots, N-1$. On the other hand, (3.41) generalizes toward "diffused" reflection: incoming waves are scattered out to all output directions, and the reflected waves are weighted by their respective directions of incidence.

We have compared the low-frequency mode positions in a 2-D triangular WM modeling a rectangular membrane, simulating both the ideally rigid and soft boundary[2]. In both cases, the substitution of the individual treatment of single boundary reflections, by means of DWFs, with the integrated processing of multiple reflections, involving the use of 2-, 3- and 4-port boundary scattering junctions (refer to Figure 3.4), led to more accurate positioning of the natural resonances of the structure.

Such positions are listed in Table 3.4 along with the mode positions provided by theory in the ideal case [36]. From that table, it can be seen that the average error (listed in the last row) is smaller when the scattered boundary model is chosen instead of the DWF termination.

In Chapter 5 we will apply the surface models presented here to the simulation of bounded pressure fields.

[2] once again, we emphasize that the "downgrade" to the 2-D case is intended for simplifying the simulations, not for studying the behavior of 2-D domains.

Mode number	Ideal position	Soft wall		Rigid wall	
		DWF	Scattering	DWF	Scattering
(0,0)	0			0	0
(1,0)	0.29			0.30	0.30
(0,1)	0.33			0.33	0.33
(1,1)	0.44	0.44	0.44	0.44	0.44
(2,0)	0.58			0.60	0.60
(2,1)	0.66	0.67	0.68	0.65	0.65
(0,2)	0.67			0.66	0.66
(1,2)	0.72	0.69	0.70	0.70	0.70
(3,0)	0.86			0.86	0.86
(2,2)	0.88	0.82	0.87	0.89	0.88
(3,1)	0.93	0.94	0.95	0.94	0.91
(0,3)	0.99			0.96	0.95
(1,3)	1.04	0.97	0.99	1.03	0.97
(3,2)	1.09	0.98	1.09	1.06	1.05
(2,3)	1.15	1.01	1.10	1.08	1.09
(4,0)	1.15			1.10	1.14
(4,1)	1.19	1.03	1.22	1.12	1.15
(3,3)	1.32	1.06	1.26	1.14	1.23
(0,4)	1.33			1.17	1.27
(4,2)	1.33	1.10	1.27	1.22	1.31
(1,4)	1.36	1.11	1.31	1.24	1.36
(2,4)	1.45	1.14	1.33	1.27	1.39
Avg. error		**0.35**	**0.03**	**0.13**	**0.03**

Table 3.4. Theoretical and calculated positions of a rectangular membrane modeled by a triangular WM, provided with both soft and rigid termination. DWF and scattering boundary terminations are compared. Average errors in the different cases are listed in the bottom row.

3.4 Minimizing dispersion: warped triangular schemes

So far we have accepted dispersion as an ineluctable numerical artifact, consequence of discretizing wave propagation. Indeed, this is true in the sense specified in § 3.2.3. In that section we have concluded that we must deal with this artifact as soon as we want to simulate a membrane using the cable network analogy, and in particular its formulation in the discrete time, that is, the rectangular WM.

At this point, we wonder whether we can alter the properties of the WM in order to come up with a "transformed" version of the model, in which the processing of digital wave signals resembles ideal wave propagation, as specified by (3.1), more closely.

Our goal is, clearly, flattening the dispersion function (from now denoted with D) or, equivalently, forcing $c(\boldsymbol{\xi})$ to assume values which are as much as possible constant and independent from $\boldsymbol{\xi}$. Such values are not supposed to be necessarily equal or close to the upper limit expressed by (3.20): any "flattened" propagation speed function will minimize dispersion.

A general solution for this problem seems impossible to achieve. In fact, a look at the dispersion function affecting the rectangular geometries (see Papers B and

C) reveals that, in those models, only certain frequency components are dispersed. Unfortunately, altering those frequencies determines variations in other undispersed components that are combinations of the dispersed frequencies. Hence any minimization procedure, even focusing specifically on the dispersed components in a rectangular WM, cannot leave other (undispersed) frequencies untouched[3].

As far as we know, an increase in the number of directions of propagation in the cable network leads to simulations that are affected by a less critical dispersion. In principle we can think that increasing this number results in more "freedom" for the (mono-dimensional) wave components to move along a specific direction. In the limit case, a cable network "looks like" a membrane, or a 3-D pressure field, as long as the cables cover all the (infinitely) possible directions. This interpretation of the propagation domain as a composition of infinitely short waveguides oriented along infinite directions is supported by the form taken by the analytical solutions provided by Equations (3.1) and (3.2). In fact those solutions integrate multidimensional spatial frequency components moving along all possible directions [102]. Huygens' principle proposes a similar model of multidimensional wave propagation as well.

Returning to our WM models, we note that triangular geometries (made of 6- and 12-port junctions respectively in the 2-D and 3-D case) are affected by a smoother dispersion function compared with rectangular (4- and 6-port junctions in 2-D and 3-D domains, respectively) WMs (again see Papers B and C). That function is sufficiently independent of the direction of wave propagation in a way that, with good approximation, in the triangular case we can reconsider dispersion as a single-variable function, depending only on the magnitude of the frequency components [137]. In other words, we can write

$$D(\boldsymbol{\xi}) \approx D\left(\sqrt{\boldsymbol{\xi}^T \boldsymbol{\xi}}\right) = D(\xi) \quad , \tag{3.43}$$

ξ being the spatial frequency magnitude. From this viewpoint, dispersion depends with good approximation only on the frequency of a spatial component, independently of the direction it propagates along the WM.

Holding (3.21), then (3.43) implies $f \approx \xi c(\xi)$: this means that if we are able to change the propagation speed of the individual spatial traveling wave components, then we can take advantage from this change to produce corresponding variations in the temporal frequency positions.

In a WM, the propagation speed is proportional to the rate at which adjacent scattering junctions exchange wave signals. In § 3.1.2 we have seen that this rate is set to be equal to one sample exchange between adjacent junctions per time step: this choice ensures explicit computability of the scheme. Any rate acceleration would result into a non-causal, hence inconsistent, scheme. On the other hand, decreasing the rate by one or more time steps, i.e., substituting the term $NT + T$ with $nT + \delta T$, $\delta > 1$ in (3.8a) and (3.8b), would just add $\delta - 1$ idle cycles in the computations, with no advantages in terms of dispersion reduction.

[3] A more general treatment of the rectangular models, in the framework of Finite Difference Schemes, leads to 2-D and 3-D rectangular grids in which dispersion affects *all* the frequency components [14]. This would perhaps allow the design of strategies for reducing dispersion [138].

What if we would be able to implement a *fractional* delay? More precisely, what if we may set $\delta = \delta(\xi)$, with δ not necessarily being an integer number?

The theory of Digital Signal Processing provides a filter block that comes useful for our needs. This block generalizes the unit delay into a 1-st order transfer function, that delays any frequency component of a specific fraction of the time step without altering their magnitudes—for this reason it is called the *allpass* block [120]. The 1-st order allpass transfer function is equal to

$$A(\tilde{z}) = \frac{\tilde{z}^{-1} - \lambda}{1 - \lambda \tilde{z}^{-1}} \quad , \tag{3.44}$$

where we have denoted the z-transformed variable with $\tilde{z}$ instead of z, without loss of generality.

We see that we have one free parameter available in that transfer function, λ, whose optimal choice (holding the stability condition $|\lambda| < 1$) is the key for reducing dispersion. In fact we will cancel this artifact, to a level limited by the accuracy with which $D(\xi)$ approximates $D(\boldsymbol{\xi})$, if a value for λ exists providing a frequency-dependent delay $\delta(\xi)$ which satisfies condition

$$\frac{\delta(\xi)}{c(\xi)} = \text{const} \quad . \tag{3.45}$$

In this way, any change in the propagation speed is compensated by a proportional variation in the (fractional) time the wave signals take to move from one junction to an adjacent one.

Unfortunately, the freedom provided by the 1-st order allpass block does not afford such an accurate control of the propagation speed. Its integration in the triangular WM nevertheless results into a major reduction of the dispersion artifact. More in general, the complications arising from choosing the allpass instead of the pure delay must be taken into account, and justify that choice only when critical constraints exist in the required model accuracy.

An optimal choice for λ can be made if we look at (3.44) as a *domain transformation* F in the z domain [101]. In that perspective, substituting z^{-1} with $A(\tilde{z})$ corresponds to mapping the z variable into the new variable $\tilde{z}$ according to:

$$z^{-1} = F^{-1}(\tilde{z}) = A(\tilde{z}) = \frac{\tilde{z}^{-1} - \lambda}{1 - \lambda \tilde{z}^{-1}} \quad . \tag{3.46}$$

Techniques based on allpass transformations of the z variable are frequently mentioned in the literature as *frequency warping* [70]. In fact, (3.46) translates into a remapping of the frequency domain of the system response. In our specific case, (3.46) can be tuned in the parameter λ in order to minimize (according to some norm) the *distance* between ideal (undispersed) and transformed (dispersed and warped) temporal frequency positions [137].

A computational complication arises after the application of the transformation (3.46). In § 3.1.2 we have seen that having at least one between-junctions delay was a necessary condition for the explicit computability of the WM. As we can imagine, the mapping of the delay into a frequency-dependent fractional delay results in the loss of explicit computability. Nevertheless, the transformation (3.46) does not

alter the stability and energy-conservation properties of the WM. For this reason, it is possible to look for a strategy that resolves the implicit calculations presented by the WM once it is transformed by (3.46).

Savioja and Välimäki have solved the computability problem offline [137]. In that solution they *dewarp* the signal that feeds the model, then process it through a triangular WM, and finally warp the output. Dewarping and warping are performed using maps equal to, or directly derived from (3.46). In this way, they bypass the loss of explicit computability in the warped WM.

Alternatively, we can look for transformations that preserve explicit computability. Those transformations must necessarily factor out a unit delay. Thus, their order will be greater than one. One possible transformation preserving explicit computability is the following one:

$$z^{-1} = F^{-1}(\tilde{z}) = \tilde{z}^{-1} A(\tilde{z}) = \tilde{z}^{-1} \frac{\tilde{z}^{-1} - \lambda}{1 - \lambda \tilde{z}^{-1}} \quad . \tag{3.47}$$

In Paper A, (3.47) is optimized for warping a triangular WM modeling 2-D wave propagation along an ideal membrane. For what has been said, that model does not cause any computational problem arising from an implicit solution of its propagation equations.

In Paper J we present a method for the computation of implicitly computable linear digital filter networks. That method results from the generalization of a technique solving the implicit computation problem in the theory of linear digital filter networks, known in the Digital Signal Processing theory as the *delay-free loop problem* [69]. Based on that generalization, we develop an online solution for the computation of the warped WM. Finally, we provide a strategy for minimizing dispersion, along with results coming from simulations of the warped triangular WM adopting that minimization.

4

Sounds from Morphing Geometries

[...] E poi la circolarità: il ritorno in un corpo, ma per un mondo diverso,
circolare, la cui unica legge ammessa è fidarsi nell'abbandono.
Un circuito è un ritmo ripetuto di questi abbandoni circolari, che ritornano
perfetti, puntuali, attesi: tra i rari luoghi in cui la fiducia pura sia ripagata.
[...] Come quella che riusciva ad esistere nell'anima del grande Omobono Tenni.
[Geminello Alvi, *Vite fuori dal mondo*. 2001.]

[...] And then circularity: the return to a body, but in a different, circular world,
whose only admissible law is to trust oneself into surrender.
A race track is a repeated rhythm of these circular surrenders, which come back
perfect, on time, attended: one of the rare places where pure faith is rewarded.
[...] Like the one that was able to exist in the soul of the great Omobono Tenni.
[Geminello Alvi, Vite fuori dal mondo. *2001.]*

This chapter deals with the somehow "intriguing" issue of sonification [84]: *Can we hear shapes?* The idea is not new: some experimental psychologists have investigated whether humans are able to recognize the shape of resonant objects by hearing, coming to interesting results [89, 92].

Apart from few cases, the issue of shape hearing has not been studied so much. One important reason for this lack of results is probably found in the existence of a fundamental weak point concerning the initial hypothesis of investigation. In fact, our hearing system identifies *sources* instead of objects from corresponding sounds [86]. During that identification task, each sound source is labeled with attributes (i.e., a car slowing down) that in general do not account for the object shape, unless the sound recalls (usually visual) clues calling for a particular shape of the recognized object or event[1].

4.1 Recognition of simple 3-D resonators

Nevertheless, the response of most resonators to an excitation adds distinct attributes to a sound, especially in the form of spectral modifications which are

[1] It is, for example, the case of a friend who is evoked *roundness* when she hears a drop falling into a quite water poll.

well-characterized by the magnitude spectra of those responses, and can be easily heard also by naïve listeners [126]. Hence, it would not be surprising that subjects recognize a resonant component in a sound, especially if the shape of the resonator is simple, i.e., a sphere or a cube.

Rocchesso has demonstrated that subjects were able to discriminate between different sizes of "ideal" cubes and spheres [127,128]. More precisely, he has added to a sound, that was rich in its original spectral content, the contribution of the most audible frequencies existing in the ideal responses taken from corresponding cubes and spheres having different volumes. As a result, he has noticed that subjects were able to discriminate by hearing which resonator, between a sphere and a cube having different sizes, was smaller (i.e., had the smaller volume).

In that case the experiment was partially ecologic: responses from virtual instead of real objects were employed. In other words the ecologic paradigm was shifted from object reproduction to object representation. Moreover several major assumptions have been made in this representation, the most important of which actually was that humans discriminate attributes of shape in the resonators from the first (low-frequency) spectral components conveyed by those resonators, even if they are idealized. In other words, a consistent acoustic *cartoonification* of the resonators has been conducted [84], and a psychophysical evaluation of some aspects of those cartoons has been attempted consequently.

Indeed, interfaces does not need faithful representations of the original sounds. As long as small cubes or large spheres are recognized by the user to take place in the scene representation, whatever they sound like, then the sound design activity can be considered successful. On the other hand it is clear that the more ecological the sound, the more effective the acoustic representation of an object is. In this sense, sonifying a sphere "for what it sounds like" would make much more sense than using, for example, a doorbell sound to display it.

Moreover, sonifying objects by adding distinctive sets of resonances to a source sound is particularly attractive since it would leave almost complete freedom of choice to the designer about the sound source to choose. In this way, a correct ecological approach to object representation matches optimally with the need to increase the bandwidth of interaction available to the user in the audio channel—this point plays a key role also in the representation of distance, addressed in Chapter 5.

One more time, the initial question comes out: can we hear shapes? The answers coming from the experiments described here represent a starting point for a subject of investigation, whose importance in the field of HCI is certainly demanding further experimentation.

4.2 Recognition of 3-D "in-between" shapes

Suppose that we can hear cubes and spheres. Then, shall we hear shapes in between those geometries? The answer in this case is: probably we don't.

Our experiment started with the design of those "in-between" shapes. We started from an interesting observation made by Rocchesso [123], who argued that the point of *pitch equalization* occurs when the cube and the sphere have the same volume.

Then, we had to find a way to migrate from "sphericity" to "cubicity". A possible path is proposed by *superquadrics*, whose analytical formulation allows an easy definition of intermediate geometries between the sphere and the cube, by simply changing one coefficient in the corresponding formula. Hence, we synthesized superquadrics having the same volume.

By those geometries we set up corresponding resonant cavities. Accurate sets of resonances from those cavities were obtained by WM simulations of the corresponding geometries.

We have noticed that, during the migration from spherical to cubic geometry, all resonances move from a starting to a final position in frequency. Moreover, during their migration most of them split into two or more new components. Those components in their turn move along frequency, going one across the other or merging together at points where their frequencies match.

During this morphing, the sound coming from the corresponding cavities changes repeatedly from "pitched" to "unpitched" character. This behavior depends on temporary alignments of some of the frequency components into a series, which possibly evokes pitch. Such alignments are not in direct relation with the geometry of the corresponding cavities, nevertheless their effects play a major role at the auditory perception level [117].

Although this investigation did not lead to correspondences between shapes and sounds, it is our opinion that the same experiment should be conducted also in the 2-D case, that is, for square and circular resonators. Compared with the 3-D case, here the lower density of resonances in the spectrum should not lead to multiple pitched and unpitched sound occurrences during a single migration of the shape from the circular to the square geometry. As a possible consequence, those sounds may exhibit a coherent timbral migration. Such an experiment could be object of further research on shape hearing.

In Paper E the details of our model for the synthesis of sounds from intermediate geometries between the sphere and the cube are given. In Paper F we also give a model for real-time synthesis of the same sounds.

5

Virtual Distance Cues

Poi, compresi. Compresi la santa, rassegnata aria da icona bizantina dei vecchi pescatori
e il selvaggio alcolismo dei più giovani. Compresi la cosa sospesa su di noi tutti,
mai ammessa, mai menzionata, ma presente, ogni giorno, dopo ogni tuffo.
Nessuno scherza con un pescatore appena risalito da una lunga immersione [...]
[Peter Throckmorton.]

And then, I understood. I understood the holy air of resignation of the old fishermen,
like that of a Byzantine icon, and the heavy drinking of the youngest.
I understood there was something hanging above us all,
never admitted, never mentioned, but present everyday, after every dive.
Nobody jokes with a fisherman who has just come up from a long immersion [...]
[Peter Throckmorton.]

3-D WMs enable to set up representations of reverberant environments, where acoustic waves travel according to the propagation model specified by (3.7) and (3.8a)/(3.8b). The numerical properties of the WM have been outlined and compared with the ideal case in Chapter 3.

In that model we have also been able to develop strategies for the simulation of locally reacting surfaces, along with rules for setting consistent absorption and reflection parameters. Those arguments are explained in Chapter 3 as well.

In this chapter, first we outline how to build a WM model of a listening environment. The reader wishing to go in more detail is referred to Paper B, where an application for simulating various configurations of WM models is described. Then, we deal with the question of how to optimize a WM model on a perceptual basis, in order to provide reliable distance cues to a listener. Once again, the reader can have further details by Paper G.

5.1 WM model of a listening scenario

The WM model we finally came up with in Chapter 3 simulates, with the desired approximation, the following features (refer to Figure 2.1):

- size and shape of the enclosure;
- position, size, shape, directivity, reflection and diffraction properties of the sound source;

- position, size, shape, reflection and diffraction properties of objects in the scene;
- position and reflection properties of walls.

Reproducing these features is straightforward:

1. Enclosure *size* must be derived by (3.20), once the propagation speed of sound (around 343 m/s in normal environmental conditions [102]) and the sampling frequency of the simulation (usually between 8 kHz and 48 kHz) have been set.
   In addition to that we have to deal with dispersion. In the simplest case we can approximate the wave propagation speed to be equal to the upper speed limit (3.20). From that we obtain the unitary spatial step size:

$$D = \sqrt{3}\, c/F_s \quad . \tag{5.1}$$

   By (5.1) we can immediately choose the number of scattering nodes, N_D, we have to put in series to reproduce a given resonator length S. This number is equal to $N_D = \text{round}(S/D)$.
   Size approximations can be reduced as long as we decrease the unitary spatial step D. This can be done by increasing F_s, although the computational cost coming from changing the sampling frequency is significant. In fact, from (5.1) it can be easily derived that any variation in F_s causes the number of junctions to change proportionally to 10^3. Moreover, digital audio signals contain a number of samples which is proportional to their sampling frequency. Finally, varying F_s changes the complexity of the model by a factor equal to 10^4.
2. The *shape* of the resonator depends on the way scattering elements are assembled together in the model. Nevertheless, it must be noticed that the WM geometry affects shape design. For example, rectangular geometries lead to object representations made of a set of cubes, however assembled. On the other hand, triangular geometries result in objects which are compositions of triangular pyramids (refer to Figure 3.4).
3. Object *positions* can be straightforwardly selected by choosing their coordinates in the 3-D discrete set of scattering nodes forming the WM.
4. Object and wall *reflections* are set by including DWFs or boundary junctions in the WM (refer to Section 3.3).
5. *Diffraction* is automatically modeled by wave scattering around WM boundaries.

Although in principle we could represent also the listener as part of the scene, and provide him with features such as the above ones, we deliberately avoided dealing with subjective cues, as told in § 2.1.5. Alternatively, a fine-grained (although computationally very expensive) WM including also a description of the listener would probably provide at least some of the attributes which are typical objects of investigation in structural modeling (see § 2.1.4).

In Paper B an application devoted to the design and execution of several WM configurations is described.

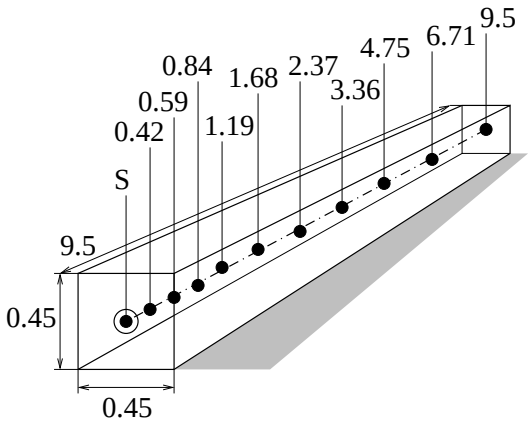

Fig. 5.1. Listening tube. A sound source (S) is heard from several points located inside the tube model.

5.2 Distance attributes

§ 2.1.3 outlines some psychophysical results in distance perception. In that section, we have learned that an ultimate strategy for synthesizing distance cues is far beyond the researchers' expectations. The bunch of possible strategies ranges from trivial solutions (loudness variations mapped by the sound source distance) to sophisticate reverberation models, providing direct-to-reverberant ratios and, possibly, lowpass filtering depending on the distance parameter.

In particular, the latter approach should account for absolute distance cues such as those described in § 2.1.3, otherwise impossible to create for unfamiliar sources located in the mid range. Unfortunately, a WM reverberation model accounting for the middle range leads to computationally expensive simulations.

Reasonable doubts (refer to § 2.2.4) exist about the need to add sophisticate attributes in the reverberant signal energy in order to evoke a sense of distance. On the other hand, (even sophisticate) artificial reverberators in general do not make distance controls available to the user.

5.3 Synthesis of virtual distance cues

In this section we propose to simulate a simple listening environment that, on an intuitive basis, cannot convey any definite spatial impression but a sense of distance. It consists of a tube, around 10 m long, having a square section approximatively 0.5 m wide (see Figure 5.1). On one of the two edges we put an unfamiliar sound source. Finally we ask a phantom listener, located somewhere inside the tube, to rate the perceived distance existing between himself and the sound source.

We realized this listening environment using a 3-D rectangular WM sampled at 8 kHz, provided with WDFs modeling the internal surfaces. We modeled an omnidirectional point-wise sound source, and two point-wise acquisition points separated by a distance approximately equal to the average interaural distance.

Variations in the listening position were simply realized by corresponding changes in the acquisition points, along a direction parallel to the longest dimension of the tube.

Changes in the direct-to-reverberant energy ratio can be applied, by varying the tube internal surface absorption parameters. More in general, other aspects of the reverberant sound can be manipulated by varying the surface impedances: such manipulations can even lead to major timbral changes in the sound. Despite this, the sense of distance provided by the model exhibits a good independence from those changes, suggesting an inherent robustness of model with respect to parametric variations. It is noticeable that robustness also holds with respect to the sampling frequency: in fact, the 8 kHz model provides reliable distance cues.

The generation of robust cues is probably a consequence of the tubular geometry of the listening environment. In practice, this environment produces exaggerated cues. For this reason, the reverberant sounds synthesized by the model could hardly find an application in musical sound processing. On the other hand, the same sounds look attractive for applications of virtual reality [9] and in speech and non-speech audio communication systems and interfaces [39].

In Paper G a broader description of the listening tube is given. Together with that description, results are discussed coming from psychophysical experiments conducted using the listening tube. Such results seem quite promising, suggesting that the listening tube, used as a "distance post-processor" of anechoic (or at least dry) unfamiliar sounds, provides attributes that result in consistent and robust distance cues especially for mid- and far-field sound source distance rendering.

Furthermore, the same listening environment has been applied to the distance rendering of synthetic sounds that should be conveyed to blind persons, to augment or substitute the acoustic scene representation they are able to get from the surrounding environment. Those sounds are intended to "label" corresponding objects in the scene—especially those which are inaudible and hardly detectable using alternative tools, such as the cane. Clearly, the information contained in those sonic labels must be organized in such a way that its extraction from the acoustic message requires the least effort for the user.

In this application the source/listener distance is a critical factor, in fact it accounts for part of the information characterizing the relative object position. Nevertheless, adding reliable distance cues to sounds in the form of reverberant energy leaves other perceptual dimensions, in particular all the dimensions which can be specified by the acoustic features of the sound source, available for further object characterization. Moreover, distance cues conveyed by reverberation are recognized naturally and straightforwardly by listeners.

6

Auditioning a Space

The door of a Cadillac sedan shuts with a velvety sound bespeaking luxury. The Vespa scooter is a Mediterranean insect which for a long time heralded spring, the 2CV a hard-wearing, rickety alarm clock. The Singer sewing machine clicks dryly and the Frigidaire closes with an opulent 'chunk' which inspires confidence. What would all these 'objects' be without their distinctive voices?
[Louis Dandrel, *The voice of things*. In *Industrial design*, Jocelyn de Noblet, Ed. 1993.]

Our sound *chain* must start with a sound synthesis block (for example the crumpling sound synthesizer we will outline in Chapter 7), and includes a spatialization block which adds distance attributes as those described in Chapter 5. Together, those blocks generate sounds according to the model specifications and the user's input parameters.

We are now dealing with the last stage of this chain. This stage is responsible for sound reproduction.

As part of the overall information displayed to a user by a human-computer interface, sound is presented through some kind of audio reproduction system. Now, the question is: how to present this information to the user? There are many ways, and many reproduction systems available for setting up an auditory display. Those systems range from sophisticate audio equipment available on board of state-of-the-art VR installations [9, 93] to tiny loudspeakers mounted on commercial portable devices, such as mobile phones and palmtops. No doubts that the presentation strategies adopted in those two extreme cases are dramatically different [31].

Nevertheless, the perceptual issues of auditory localization that we have briefly introduced in Section 2.1 hold independently of the display method. Hence, for a given sound spatialization model we must look for a strategy which optimizes the presentation of that model on the psychophysical requirements of human listeners. Though, we have to take many practical factors into account during the choice of that strategy.

In the following of this chapter we will look for an optimal presentation method for our spatialization model. Our optimization criterion is manifold. In fact, during the choice of the reproduction system we will consider the following three aspects:

- quality of the VR cues conveyed by the auditory display;
- cost of the presentation;
- usability of the interface in everyday-listening contexts.

It is our opinion that the potential technological impact of a spatialization model underlies those aspects. More in general, we can never assess the performance of a sound spatialization model completely, unless we are not fully aware of the way and the place where sound is presented to the listener[1]. A general discussion upon those three aspects is perhaps the best starting point to address specific application contexts where our tubular model can be used.

In particular, the third aspect is not so straightforward to deal with when the auditory mode is added to a machine interface. For instance, the presentation of sophisticate audio VR cues usually overlaps with the user's possible need to communicate with other people, especially in cooperative contexts such as those normally experienced by groups working in the same physical environment. In that case, audio messages coming from different systems, mixing together in the same environment, can degrade the performance of a workgroup instead of improving it.

If we consider today's available reproduction systems, we come up with two basic solutions for the presentation of sounds: loudspeakers and headphones.

- Loudspeaker arrangements can provide convincing cues of audio VR, especially if their action focuses on a single user [57, 83]. On one hand they enable the user to interact with other persons joining the same environment using natural communication. On the other hand sounds displayed using loudspeakers become distracting for the rest of the people as soon as they are engaged in other tasks: those sounds are in fact perceived such as external interferences.
- Headphones guarantee the mutual independence of different auditory displays. Unfortunately, they also represent an obstacle for natural communication, as they prevent the user to talk directly with the rest of the group.

We have conducted experiments with our spatialization model using both presentation methods. In both cases we avoided to use specific equipment and subjective binauralization strategies. Both the loudspeaker and headphone presentations nevertheless conveyed convincing distance cues to listeners. In the following, we will understand why.

6.1 Open headphone arrangements

A possible solution to the problem of audio presentation perhaps lies somehow in between headphones and loudspeakers. We have surveyed several existing headphone designs, to investigate the eligibility of *open* headphones as optimal reproduction systems for our spatialization model. This analysis is reported in detail in

[1] This is true to an even broader extent. To pick up an example coming from musical instrument design, in that field it is true that sound designers never specify the timbre of an electronic instrument before having an idea about the characteristics of the listening context where that instrument will be mainly played [63].

Paper I, which results from gathering a certain amount of literature (some of it not being immediately available) existing in the subject of headphone reproduction. This paper comes out from an effort to point out the state of the art in this subject.

It seems that the open ear-phone design will represent, in the near future, a low-expensive way to achieve convincing audio presentations, with low or no interference with the information streams which are usually exchanged among individuals working together in the same physical environment.

More precisely, returning back to the three points addressed in the beginning of this chapter:

- effective audio VR cues can be conveyed as long as specific distortions introduced by headphones are compensated. Those distortions originate due to the proximity of the headphone audio transducers to the listener's ears, and due to the unpredictability of the transfer characteristic of the *pressure chamber* formed by sealing the external ear with the headphone cushion. The macroscopic perceptual effect of such distortions is known as *Inside-the-Head Localization* (IHL) [15], usually coming together with other effects [131].
  It is likely that the open headphone design strongly reduces or even eliminates the pressure chamber effect. Moreover, open headphones allow external noises to mix with the audio message displayed by the interface, this mixing having major benefits in decreasing the perceived vicinity of the headphone transducers;
- headphones cost less than loudspeaker equipment having comparable figures of sensitivity, signal-to-noise rejection and so on;
- headphones virtually eliminate interferences caused by simultaneously displaying independent audio messages in the same physical environment. Moreover the open design allows the user to hear also the external sounds, and to interact with the rest of the world using natural communication and avoiding unusual vocal intonations during talking, caused by partial deafness against his own voice.

One more thing must be considered when sounds are reproduced using headphones. Although in principle the open design can enable audio VR, nevertheless headphones cannot convey cues of absolute localization: since they are firmly imposed over the listener's head, then the represented position in space of the sound source shifts together with head movement. For this reason, audio VR presentations through headphones should include a system for head-tracking, in a way that *dynamic* cues can be conveyed to the listener based on his head position [57]. At that point, the interface is enabled to represent absolute sound source positions through the auditory channel.

6.2 Audio Virtual Reality

Going into some more detail, for what we have seen in Chapter 5, Section 2.1 and Section 6.1 the synthesis of audio VR cues requires the design of a model that must contain the following blocks in sequence:

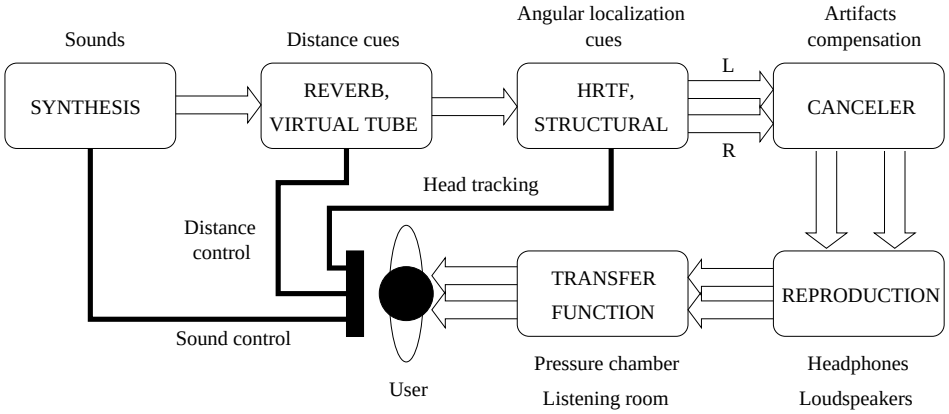

Fig. 6.1. Model for the generation and presentation of audio VR cues, including distance. Sounds are provided with distance cues, then spatialized with HRTF or structural models to add angular position cues. Before presentation to the user, the resulting sounds are processed by a canceler which compensates artifacts coming from reproduction equipment and transducer-to-ear transfer functions. The user, in his turn, interacts with the sound synthesis and sound spatialization and binauralization blocks.

a sound synthesis block – this block reproduces the characteristics of the sound source. We will briefly deal with an example of physics-based sound synthesis in Chapter 7;

a spatialization/binauralization block – such a block adds spatial attributes, providing distance and angular localization cues such as those outlined in § 2.1.1, § 2.1.2 and § 2.1.3;

a cancellation block – this block is devoted to compensate all the undesired attributes affecting the sound at the presentation. More precisely, any artifact added to the sound by the reproduction chain, that is, distortions introduced by the audio equipment plus transducer-to-ear transfer functions, depending also on the characteristics of the *real* listening environment [104, 105], should be canceled in a way that the sound presented at the ear contains, with very good approximation, only the attributes provided by the synthesis and spatialization model, instead of a distorted version of those attributes due to the presence of the aforementioned artifacts.

The block diagram of a model for audio VR that provides also distance rendering is depicted in Figure 6.1.

So far our research has only dealt with the sound spatialization block. Nevertheless, for the sake of completeness we have also analyzed some aspects of the canceling block. Taking for granted that convincing binauralization techniques can be found in the literature (refer to § 2.1.1 and § 2.1.2), i.e., that satisfactory angular localization cues can be reproduced, then we want to gain a certain control over the last block where digital sound processing can be applied; that is, the compensation block.

6.2.1 Compensation of spectral artifacts

The literature about compensation of audio artifacts is huge. Depending on the type of artifact, corresponding techniques have been found for recovering the original characteristics of a sound. An outline of the various existing methods aiming at restoring the audio information from a signal corrupted by the presence of some artifact is beyond the scope of this work, and can be found (often limited to a specific branch of audio restoration) elsewhere [64].

Here we focus on one of the most popular compensation techniques known in the field: *parametric equalization* [20, 98]. Despite their versatility, parametric equalizers allow a precise manipulation of the magnitude spectrum of a sound. By means of this manipulation, even subtle aspects in the spectrum of a sound reproducing a particular acoustic space can be enhanced. Conversely, resonance frequencies can be smoothed when they are excessively prominent in the same spectrum (for example when the sound is acquired in points of the acoustic space where some of them are too evident): in the latter case the equalizer is used as a canceler. Though, it must be reminded that parametric equalizers cannot cancel non-minimum phase components in the acoustic signal [110].

Compensation of spectral artifacts basically requires to measure the response of the acoustic space over one or more listening points [17, 121]. Once the spectral characteristics of those responses are figured out, a network of filters is tuned in a way that spectral attributes *reciprocal* to the measured ones are added to any sound prior to its audition. This network can be created using a number of parametric equalizers in series (one equalizer per frequency component, in the best case).

Parametric equalizers cannot be straightforwardly reciprocated. In other words, determining the inverse transfer characteristic of a parametric equalizer is not immediate. Although it is easy to show the reciprocal equalizer exists, nevertheless its design moves through a patient algebraic rearrangement of the original transfer characteristic.

This algebra can be fairly exciting to deal with, nevertheless it is still rewarding as long as it is worked out offline in a way that the inverse filter coefficients are written in the memory of the processing device prior to the execution of the application. In fact their online computation is excessively time-consuming and hard to do in finite-precision arithmetic, such as that available in a general purpose DSP [103]. For this reason a huge amount of memory of the device is usually reserved to record the table look-ups containing the inverse coefficients as long as a reciprocal equalizer is implemented.

Inverse equalizers become particularly interesting as long as parametric equalization is employed for canceling purposes [107]. Consistently with our philosophy of keeping the driving parameters of our models explicitly available as direct controls of the model functionalities, we have designed a novel parametric equalization filter structure whose coefficients are linked to the inverse equalization parameters more directly than traditional parametric structures [120]. Although our filter demands slightly more computations per time cycle, on the other hand its coefficients need simple algebraic rearrangement as long as the characteristics that we want to cancel in a transfer function vary.

In Paper H the design of our equalization filter is addressed.

6.3 Presentation of distance cues

Although we have seen that effective techniques exist for synthesizing binaural sounds, and for preventing them from being corrupted by most of the artifacts arising during the presentation, nevertheless it is true that audio VR, for its many requirements and overall cost, is still confined to specific contexts.

The question now is: does our tubular model make sense even in the case when audio VR is not guaranteed? In other words, is the distance rendering block depicted in Figure 6.1 effective even if we cut off the angular cues generation and the artifacts compensation?

An answer to that question need, first, to specify which kind of *presence* the user expects to experience [72]. To find an analogy with the visual mode, in that case it is true that a subject put in front of a conventional computer visual interface (say, a screen) expects to discriminate nearby from far displayed objects in a context where distance is represented rather than reproduced. Despite the limits of the screen interface, this discrimination can be even accurate if proper display strategies are put into action for representing distance.

On the other hand the same subject, if experiencing a more immersive virtual environment, will expect to perceive distance with more realism, perhaps the same realism he experiences in the everyday life, i.e., when stereoscopic vision is enabled. Otherwise he will rate the virtual experience as fair.

The same expectation probably have those subjects who, in front of a machine-to-user interface using conventional equipment for audio presentation, are asked to rate distance cues: as long as binaural listening is not enabled, they cannot hear but a *representation* of the auditory distance.

Once this distinction has been made clear, we still wonder if represented auditory distance works as accurately as real or realistic (i.e., using audio VR) distance. The answer is positive. This conclusion is confirmed by accurate psychophysical experiments, in which subjects evaluated relative distance by sound sources represented using loudspeakers [168]. In that case subjects demonstrated capabilities of evaluating auditory distance comparable with the performances they exhibited during experiments conducted in real listening contexts, that is, using real distant sound sources..

Our experiments using the tubular model moved a step further: since subjects wore headphones or were put in front of loudspeakers, they were aware of listening to sound source representations; in the meantime they were not told anything about the displayed scene previously to the listening test. Hence, the results we have obtained so far would confirm that satisfactory distance evaluation holds also in the case when the representation of a real scene is substituted with the representation of a virtual scene using the model described in Section 5.

Such results are summarized in Paper K, where an extensive description of the virtual tube as a display tool for multimodal human-computer interfaces is addressed.

7

Auxiliary work – Example of Physics-Based Synthesis

Superfluo, al di fuori della filologia, il commento.
Tutto Qohélet è assiomi, e assiomi *credibili*,
che chiunque può accogliere. Che cosa c'è da spiegare
in un'arpa eolica o in una emissione di muggiti?
[*Qohélet*. Introduzione di Guido Ceronetti. 2002.]

A superfluous comment, besides philology.
Qohélet is all axioms, believable axioms,
that anyone can assent. What is there to explain
in an Aeolian harp or in the emission of lowing?
*[*Qohélet. *Introduction by Guido Ceronetti. 2002.]*

We conclude the work presenting an example of physics-based sound synthesis, developed by the author as part of his research activity made during this year for the EU Project IST-2000-25287 called *SOb – The Sounding Object*. This chapter deals with the synthesis of so-called *crumpling* sounds. Such sounds occur whenever the perceived acoustic event identifies a source whose emission, for any reason, is interpreted as a superposition of microscopic crumpling events, or *impulses*.

Although physics-based sound synthesis is beyond the scope of this work, for the sake of completeness we wanted to test our spatialization model using an autonomous sound synthesizer, such as the one presented here in the following. This chapter explains how this synthesizer works. Though auxiliary, it contains interesting information which comes useful especially for a thorough comprehension of the sound examples prepared by the author using synthetic crushing and walking sounds. Those sounds are available at the author's web site.

Moreover, there is a common framework inspiring our approach both to sound synthesis and sound spatialization. As told in the Introduction, this framework asks for defining a direct link from the machine to the user and vice versa. The synthesizer proposed here follows this general principle.

Our aim is to provide models in which we can control objective parameters directly. In crumpling sounds the fundamental quantities which must be under control are:

- the structure of each *impulse*;
- the statistical distribution in time and amplitude of those impulses, forming individual *events*;

- the overall temporal statistics of the event occurrences.

Each one of those elements plays a precise role in the definition of a sound sequence [34]. In particular, we have noticed that if we keep a unique structure for all impulses, and we add a notion of *energy* associated to the amplitude of each impulse, then we can control quantities such as object *size*, applied *force*, and material *stiffness* directly, via the tuning of explicit parameters contained in the statistical laws governing the temporal and magnitude evolution of the impulse sequences. Those statistical laws can be figured out starting from natural considerations about the event dynamics.

Examples of sounds which we have found to be effectively synthesized using crumpling events are crushing cans. Examples of sound sequences resulting by adding a higher-level statistics to crumpling events are footsteps.

Our work has mainly focused on the structure and statistics of the impulses. Extensive research on higher-level structures of complex sound sequences can be found elsewhere [19]. *Object-based* sound synthesis relies on generation models that privilege easy access to the control layer of the sound synthesis algorithm. This approach turns out to be particularly effective in the *sonification* of dynamic interactions between objects, where sounds from ecological events such as rolling, crumpling, bouncing, are reproduced [6, 34].

The perceived quality of such sounds does not depend mainly on the accuracy in the synthesis. Rather, their coherence with respect to an event plays a major role in increasing sound consistency, better if this event is represented to the listener through alternative information such as visual display of the same event or evidence about the existence of the cause determining it.

This correspondence between everyday events and sounds is supported by *ecological* experiments where the recognition of events by ear was tested avoiding the use of pre-recorded sounds, instead using proper audio representations (or "cartoon" models) of those events [165]. The modification, in the "cartoon" models, of parameters that are related to the event dynamics resulted to be more relevant, in those experiments, than other factors such as spectral similarities between "cartoon" sounds and faithful recordings of real events.

The object-based approach provides models allowing a precise and fast access to the physical quantities that determine the event "nature". Although this approach translates in lower sound accuracy compared with audio sampling, nevertheless object-based modeling appears to fit particularly well with the ecological approach to sound event perception.

7.1 Crumpling cans

Aluminum cans emit a characteristic sound when they are squashed by a human foot which, for example, compresses them along the main axis of the cylinder. This sound is the result of a composition of single "crumpling" events, each one of those occurring when, after the limit of bending resistance, one piece of the surface forming the cylinder splits into two facets as a consequence of the force applied to the can.

The exact nature of a single crumpling event depends on the local conditions the surface is subjected to when folding occurs between two facets. In particular, the types of vibrations that are produced are influenced by shape, area, and neighborhood of each facet. Moreover, other factors play a role during the generation of the sound, such as volume and shape of the can. The can, in its turn, acts as a volume-varying resonator during the crumpling process.

A precise assessment of all the physical factors determining the sound which is produced by a single crumpling event is beyond the scope of this work. Moreover, there are not many studies available in the literature outlining a consistent physical background for this kind of problems. Studies conducted on the acoustic emission from wrapped sheets of paper [73] concluded that crumpling events do not determine *avalanches* [143], so that fractal models in principle cannot be used to synthesize crumpling sounds [139]. Nevertheless, crumpling paper emits sounds in the form of a stationary process made of single impulses, whose individual energy E can be described by the following power law:

$$P(E) = E^{\gamma} \quad , \tag{7.1}$$

where γ has been experimentally determined to be in between -1.3 and -1.6. On the other hand a precise dynamic range of the impulses is not given, although the energy decay of each single impulse has been found out to be exponential.

Another, perhaps the most important factor determining the perceptual nature of the crumpling process resides in the temporal patterns defined by the events. A wide class of stationary temporal sequences can be modeled as *Poisson processes*. According to them, each time gap τ between two subsequent events in a temporal process is described by an exponential random variable with *density* $\lambda > 0$ [111]:

$$P(\tau) = \lambda e^{\lambda \tau} \text{ with } \tau \geq 0 \quad . \tag{7.2}$$

Assuming a time step equal to T, then we simply map the time gap over a value kT defined in the discrete-time domain:

$$k = \text{round}(\tau / T) \quad , \tag{7.3}$$

where round$(\cdot)$ gives the closest integer to its argument value.

The crumpling process consumes energy during its evolution. This energy is provided by the agent that crushes the can. The process terminates when the transfer of energy does not take place any longer, i.e., when a *reference energy*, E_{tot}, has been spent independently by each one of the impulses forming the event s_{tot}:

$$s_{\text{tot}}[nT] = \sum_i E_i s[nT - k_i T] \quad \text{with} \quad E_{\text{tot}} = \sum_i E_i \quad , \tag{7.4}$$

where $s(nT)$ is a short discrete-time signal having unitary energy, accounting for each single impulse.

7.1.1 Synthesis of a single event

So far, we have (yet not completely) outlined one possible description of the crumpling process in terms of its energy E_{tot}, arrival times $k_i T$ and structure s of each impulse. We now deal with the synthesis of a single event.

As long as a fracture occurs between two facets, force waves spread from the new crease along those facets, and acoustic energy is produced. Here we suppose that the sound coming from each facet becomes as higher in pitch as smaller the facet is, and as louder in intensity as larger the facet is. Although, as mentioned above, each individual impulse is supposed to produce a peculiar sound, we assume that the weighted superposition of only two "prototype" sounds accounts for all impulses, once their pitch and loudness have been set according to the considerations made in the following.

This prototype sound has empirically been obtained via *modal synthesis* [1]. Seven phase-aligned frequency components are modulated in amplitude, in order to provide for each one of them a steep attack (1 ms to reach maximum amplitude) and an exponential decay controlled by the *halving time* parameter α_i, i.e.,

$$\alpha_i \ : \ e^{-\alpha_i t_{1/2,i}} = \frac{1}{2} \quad . \tag{7.5}$$

By this definition, an exponential decay weighting function is calculated, whose value halves after a time equal to $t_{1/2,i}$.

The resulting sample is 30 ms long. Its parameters, e.g., the frequencies $f_1, \ldots, f_7$, amplitudes $A_1, \ldots, A_7$, and decay times $t_{1/2,1}, \ldots, t_{1/2,7}$, are summarized in Table 7.1 for each component. In that case, the prototype sound is called with a *driving frequency* f_0 equal to one. Otherwise, each frequency component f_i is multiplied by the driving frequency, thus varying the height (or *pitch*) of the sample.

i	f_i (Hz)	A_i (dB)	$t_{1/2,i}$ (s)
1	300	-20	.002
2	1100	-8	.001
3	1200	-8	.001
4	3000	-8	.001
5	4100	-8	.0025
6	4950	-8	.0035
7	5200	-8	.0035

Table 7.1. Modal synthesis parameters for the prototype sound (driving frequency equal to one).

The amplitude of the prototype signal is finally weighted by a *driving amplitude* A_0, in such a way that it slightly overflows the range of values between -1 and $+1$ during the attack: a clip of this signal to values between -1 and $+1$, compliant with the fixed-point arithmetic of the audio device, results in a non-linearity in the signal adding a "metallic" characteristic to the sound. As long as the driving frequency is increased, that factor is reduced in such a way that higher-pitched samples are prevented to sound too "metallic". The value of A_0 is shown in Table 7.2 as a function of f_0.

After clipping, the prototype sample finally takes the following form:

$$x_{f_0}[nT] = \mathrm{sign}(\tilde{x}_{f_0}[nT]) \cdot \min\{|\tilde{x}_{f_0}[nT]|, 1\} \quad , \tag{7.6}$$

f_0	A_0 (dB)
$f_0 < 1.2$	16
$1.2 \leq f_0 < 1.4$	15
$1.4 \leq f_0 < 1.6$	14
$1.6 \leq f_0 < 1.8$	13
$1.8 \leq f_0 < 2$	12
$2 \leq f_0 < 3$	8
$f_0 > 3$	6

Table 7.2. Amplitude weighting factor as a function of the driving frequency.

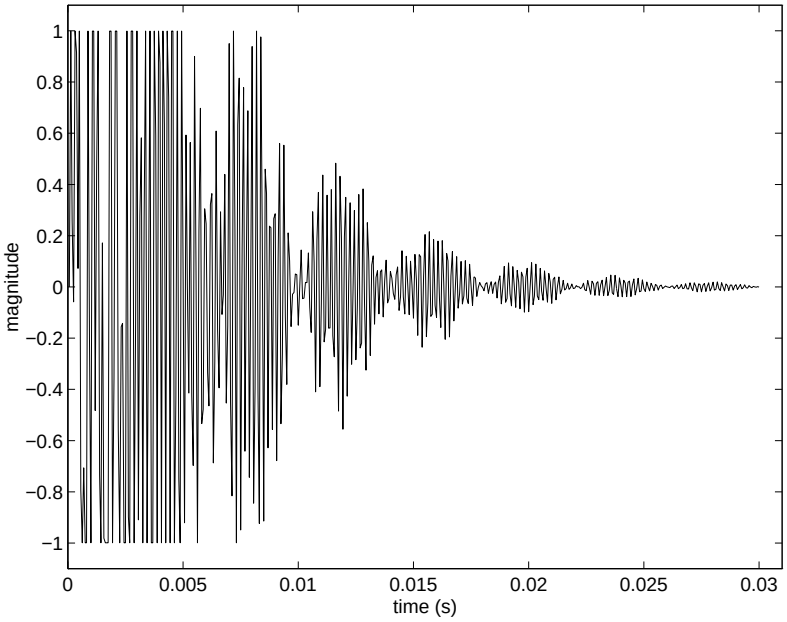

Fig. 7.1. Plot of the prototype sound ($f_0 = 1$).

where sign$(\cdot)$ gives the sign of its argument, and it is

$$\tilde{x}_{f_0}[nT] = a[nT] \sum_{i=1}^{7} A_0 A_i \sin[2\pi f_0 f_i nT] e^{-\alpha_i nT} \tag{7.7}$$

where $a[nT]$ gives the aforementioned steepness to the signal during the attack phase. The corresponding signal has been plotted in Figure 7.1 for a driving frequency $f_0 = 1$. Components' phase alignment is evident from the time modulation which appears in the decay part of the signal. Such an alignment does not result in audible artifacts.

7.1.2 Synthesis of the crumpling sound

At this point, the individual components of the process and their dynamic and temporal statistics have been decided. Yet, the dynamic range must be determined.

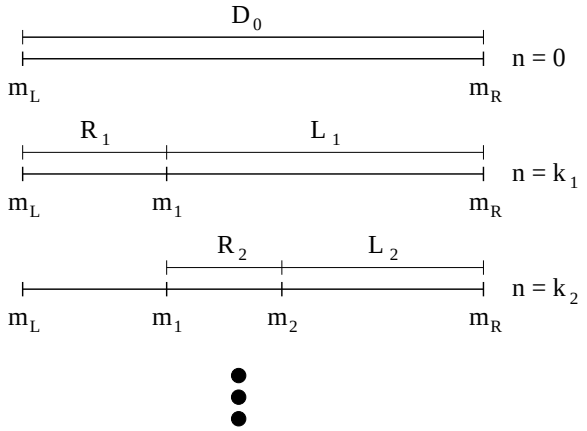

Fig. 7.2. Sketch of the procedure used to calculate the pitch of the impulses as long as the process evolves.

Suppose to constrain the dynamic range of the process to (m, M). The probability P that an individual impulse falls in that range is, using the power law expressed by (7.1):

$$P[m \leq E < M] = \int_m^M E^\gamma dE = 1 \quad . \tag{7.8}$$

This equation allows to calculate an explicit value for m if we set M to be, for example, the value corresponding to full-scale, beyond which the signal would clip. In this case we find out the minimum value coping with (7.8):

$$m = \{M^{\gamma+1} - \gamma - 1\}^{\frac{1}{\gamma+1}} \quad . \tag{7.9}$$

We still have to determine a rule for calculating the driving frequency each time an impulse is triggered.

During the crushing action over the can, creases become more and more dense over the can surface. Hence, vibrations over the facets increase in pitch since they are bounded within areas that become progressively smaller. This hypothesis inspires the model that is used here to determine the pitch related to an individual impulse.

Let us consider a segment having a nominal length D_0, initially marked at the two ends. Let us start the following procedure: Each time a new impulse is triggered, a point of this segment is randomly selected and marked. Then, two distances are measured between the position of this mark and its nearest (previously) marked points. The procedure is sketched in Figure 7.2, and it is repeated until some energy, as expressed by (7.4), is left to the process.

The values L_i and R_i, corresponding to the distances calculated between the new mark m_i (occurring at time step k_i) and the leftward and rightward nearest marks (occurred at previous time steps), respectively, are used as absolute values for the calculation of two driving frequencies, f_L and f_R, and also as relative

weights for sharing the energy E_i between the two prototype sounds forming the impulse:

$$E_i s[nT - k_i T] = E_i \frac{L_i}{L_i + R_i} x_{f_L}[nT - k_i T] + E_i \frac{R_i}{L_i + R_i} x_{f_R}[nT - k_i T] \quad , \quad (7.10)$$

where the driving frequencies are in between two extreme values, f_{MAX} and f_{MIN}, corresponding to the driving frequencies selected for a full and a minimum portion of the segment, respectively:

$$f_L = f_{\mathrm{MAX}} - \frac{L_i}{D_0}(f_{\mathrm{MAX}} - f_{\mathrm{MIN}})$$
$$f_R = f_{\mathrm{MAX}} - \frac{R_i}{D_0}(f_{\mathrm{MAX}} - f_{\mathrm{MIN}}) \quad .$$

7.1.3 Parameterization

Several parameter configurations have been tested during the tuning of the model. It has been noticed that some of the parameters outlined above have a clear (although informal) *direct* interpretation:

- E_{tot} can be seen as an "image" of the *size*, i.e., the height of the cylinder forming the can. This sounds quite obvious, since E_{tot} governs the time length of the process, and this length can be in turn reconducted to the can size. Sizes which are compatible with a natural duration of the process correspond to potential energies ranging between 0.001 and 0.1;
- low absolute values of γ result in more regular realizations of the exponential random variable, whereas high absolute values of the exponential statistically produce more peaks in the event dynamics. Hence, γ can be seen as a control of *force* acting over the can. This means that for values around -1.5 the can seems to be heavily crushed, whereas values around -1.15 evoke a softer crushing action. Thus, γ has been set to range between -1.15 and -1.5;
- "soft" alloys forming the can can be bent more easily than stiff alloys: Holding the same crushing force, a can made of soft, bendable material should shrink in fewer seconds. For this reason, the parameter p_s governing the frequency of the impulses in the Poisson process can be related to the material stiffness: the higher p_s, the softer the material. *Softness* has been set to range between 0.001 (stiff can) and 0.05 (soft can).

Also, notice that variations in the formulation of the prototype signal, x_{f_0}, usually lead to major differences in the final sound[1]. In particular we are confident in that the statistical treatment and parameterization outlined so far is general enough for the acoustical rendering of a wide class of events, including, for example, paper crumpling and plastic bottle crushing, once the prototype sound is properly shaped to accommodate for different individual events.

[1] the definition of x_{f_0} is still prone to further improvements.

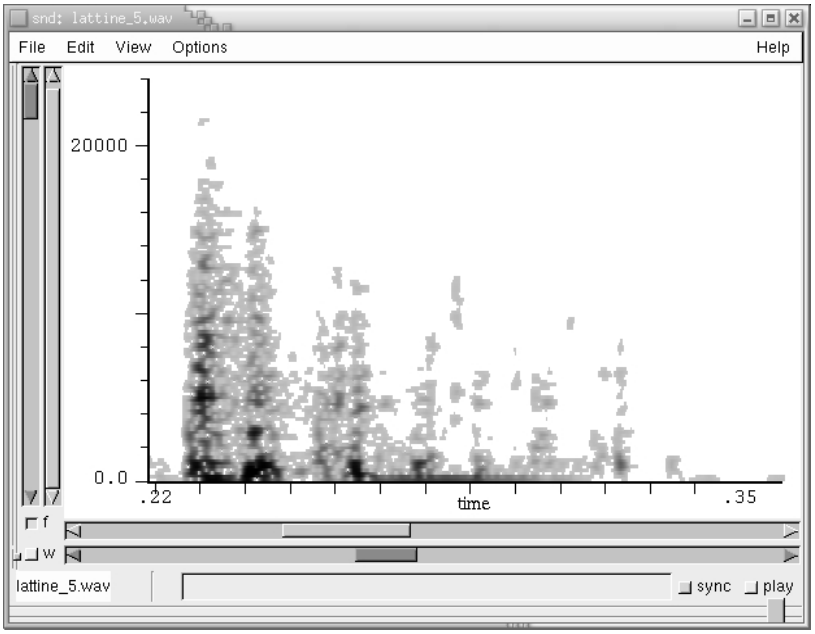

Fig. 7.3. Sonogram of the prototype sound of a crushing can.

7.1.4 Sound emission

Crushing occurs in consequence of some force acting on the can. This action is usually performed by an agent having approximately the same size as the can surface, such as the sole of a shoe.

As the agent compresses the can, sound emission to the surrounding environment changes since the active emitting surface of the can is shrinking, and some of the creases become open fractures in the surface. Moreover, we suppose that the internal pressure in the can is maximum in the beginning of the crushing process, then relaxes to the atmospheric value as long as the process evolves, due to pressure leaks from the holes appearing in the surface, and due to the decreasing velocity of the crushing action. This processes, if any[2], have a clear effect on the evolution in time of the spectral energy: high frequencies are gradually spoiled of their spectral content, as it can be easily seen from Figure 7.3 where the sonogram of a real can during crushing has been plotted.

The whole process is interpreted in our model as a time-varying resonating activity (provided by the shrinking can), simply realized in our model through the use of a low-selectivity linear filter whose lowpass action over the sound s_{tot} is slid toward the low-frequency as long as the process evolves.

Lowpass filtering is performed using a first-order lowpass filter [98]. For this particular case we adopted the following filter parameters:

- lowest cutoff frequency $\Omega_{\mathrm{MIN}} = 500$ Hz
- highest cutoff frequency $\Omega_{\mathrm{MAX}} = 1400$ Hz.

[2] We are still looking for a thorough explanation of what happens during crushing.

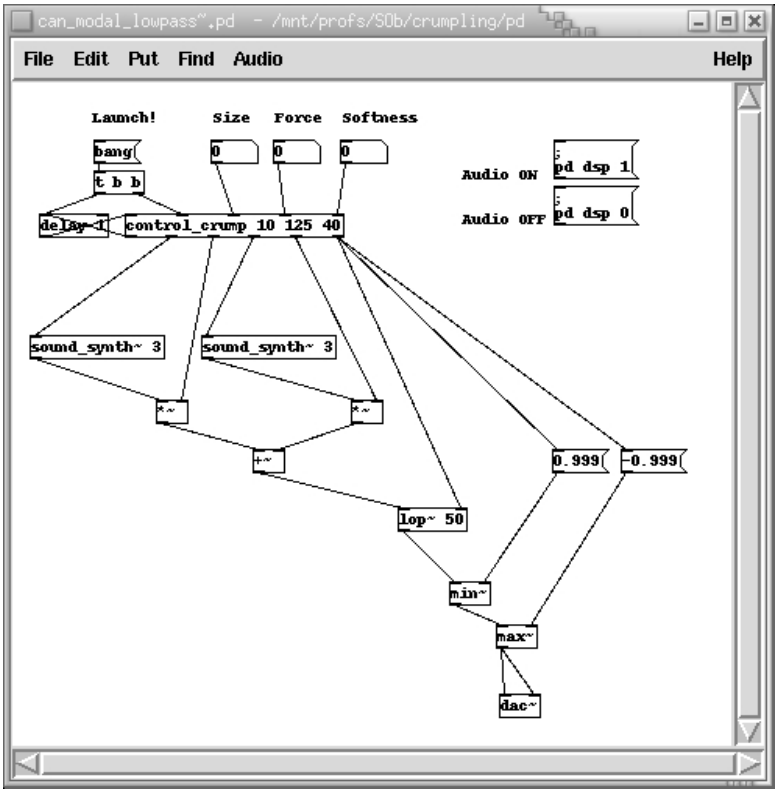

Fig. 7.4. Screenshot of the `pd`-module implementing the crumpling can model.

Using those parameters, the cutoff frequency is slid toward the lowest value as long as energy is spent by the process. More precisely, the cut frequency ω_i at time step k_i is calculated according to the following rule:

$$\omega_i = \Omega_{\text{MIN}} + \frac{E_{\text{tot}} - \sum_{k=1}^{i} E_k}{E_{\text{tot}}} \left(\Omega_{\text{MAX}} - \Omega_{\text{MIN}}\right) \tag{7.11}$$

This kind of post-processing contributes to give a smooth, progressively "closing" characteristic to the crumpling sound.

7.2 Implementation as *pd* patch

The model has been finally implemented as a `pd` *patch* [114]. This software allows a straightforward implementation, capable of maintaining the higher-lever statistics, the sound synthesis and the post-processing models decoupled inside the patch (see Figure 7.4). The only limitation with this implementation is represented by the firing rate with which the statistical module (labeled as `control_crump`) can feed control data to the two sound synthesis modules (labeled as `sound_synth~`) producing the signals x_{f_L} and x_{f_R}, respectively. This limitation comes from the

presence of the explicit feedback structure, containing the statistical module in loop-back with a `delay` block. This limits the shortest firing rate to 1 ms.

On the other hand, the chosen implementation allows an independent design of the statistical and the sound synthesis models. More precisely, the former has been realized in C language, as a `pd` class, whereas the latter has been implemented as a sub-patch, taking advantage of the several pre-existing building blocks provided with `pd`.

This modular approach leaves the patch open to changes of the individual modules. This enabled us to substitute the `sound_synth~` modules with an alternative model, which simulates the non-linear impact between two objects [118]. By properly tuning the impact parameters in that model, we could obtain even better simulations, not limited to crushing cans. More precisely convincing simulations of footsteps over a field covered with snow have been produced (refer to profs.sci.univr.it/∼fontana/) for corresponding sound samples available on the web.

Finally, footsteps events have been integrated with a higher-level event statistics to simulate walking and running over the snow [19].

Again, refer to profs.sci.univr.it/∼fontana/ for the sound samples.

Conclusion

Un architetto dovrebbe vivere 150 anni:
50 per imparare, 50 per progettare, 50 per insegnare.
[Renzo Piano. Intervista. 2001]

An architect should live for 150 years:
50 to learn, 50 to design, 50 to teach.
[Renzo Piano. Interview. 2001]

This work has proposed a design paradigm that might be useful in the modeling and representation of acoustic environments. During the conception of this paradigm we were constantly inspired by the following principle: establishing a direct relationship between the models, whatever the represented acoustic scene is, and the subjects—possibly users of a human-computer interface or active spectators in a VR scene.

This relationship is particularly effective if it is two-way, i.e., if it establishes a link both from the machine to the user and from the user to the machine. This assumption implies that any model must underlie the requirements we have outlined in Section 2.3, summarized here once again by the following two principles:

- the model must be controlled by the user directly;
- the scene must be displayed to the user directly.

In other words, we have looked for models capable of displaying scenes that did not ask the user to attend previous learning tasks, and, whenever possible, saved him from using higher cognitive stages to interpret those scenes. Meanwhile we have looked for models, whose control did not required layering intermediate maps between the interface and the model.

Clearly, those models were not at hand. Solutions providing direct access to the driving parameters had to be found in what we called physics-based models. In particular, the spatialization models we have proposed belong to the world of multidimensional distributed numerical schemes. Such schemes in any case result in computationally heavier simulations and more sophisticate realizations, compared to solutions using more traditional (i.e., monodimensional and linear) audio processing.

Once physics-based modeling has been addressed, at least we can take for granted that the first point can be worked out. Indeed, this is a consequence of the deep knowledge we have of the mechanisms that link physical quantities to the evolution of a corresponding physical phenomenon. Along with this knowledge, a lot of mathematical and numerical techniques have been developed to describe and, later, model those mechanisms into relationships between parameter settings and continuous and discrete system model behaviors.

On the other hand, the design of models which are capable of reproducing perceptually reliable sonifications and spatializations of objects and events in the scene is an even more complicate matter to deal with. In this case the lack of results explaining how we perceive those objects and events makes the design activity of virtual sound sources and spatialization contexts somehow still pioneering. Nor, as we have seen in Section 2.2, the use of traditional perceptual spatial scales such as those described by Beranek or Warusfel et al. is so helpful as soon as the ecological point of view comes into play.

This means that if, on one hand, we can resolve the human-to-machine communication directly, on the other hand we are not yet fully aware of the maps that link the characteristics of a listening environment with our perception and consequent mental representation of it.

Researchers in psychoacoustics have nevertheless made clear that our hearing system basically figures out simplified representations of the listening environment, whose spatial characteristics are detected only approximately. Starting from this assumption we have hypothesized that, similarly to what has been done in the field of physically-informed sound source design, the missing link between direct model control and direct scene representation must be filled up by simplifying the physics-based spatialization model. In this way we hoped that our models could convey exhaustive scene representations, meanwhile preserving the objective meaningfulness of the parameters.

Simplifying the models has also the fundamental advantage of reducing their computational impact, otherwise hard to deal with even in the case of offline model implementations.

This paradigm has been applied with success to distance rendering. In that case we have designed a simple, *ad hoc* virtual listening environment specialized to the synthesis of distance cues. Subjects who evaluated those cues showed promising performances during the experiments.

As part of our investigation we have introduced several improvements to the numerical scheme which we have extensively used to realize our spatialization models, that is, the Waveguide Mesh. Those improvements ranged from the definition of methods for the assessment of specific figures concerning the performance of this scheme, which were previously unknown, to the development of techniques for reducing its numerical artifacts. We think we have contributed to explain several aspects of the Waveguide Mesh, most of them having practical consequences when such schemes are used in a simulation.

By understanding those aspects, the sound designer will hopefully no longer treat the model as a kind of "black box", whose apparent behavioral idiosyncrasies often used to be resolved by increasing the sampling frequency or, equivalently, by neglecting the information coming from the high-frequency resonances synthesized

by the mesh. Such an approach is particularly inefficient when multidimensional numerical schemes are used in the simulations, since any waste of information provided by those models translates into a major loss of computational resources.

Summary and final remarks

The novelties contained in this work can be summarized in the following points:

- a spatial and temporal analysis of the Waveguide Mesh has been addressed. As a result we calculated the bandwidth of the impulse response derived from several waveguide meshes, including those having node gaps in the mesh grid;
- a modified rectangular Waveguide Mesh has been designed requiring half computational resources compared with the original rectangular scheme, meanwhile providing responses containing the same amount of information;
- a physical interpretation of the dispersion error affecting the Waveguide Mesh has been proposed starting from results coming from the theory of *cable networks*. As a consequence, the symmetry existing in the frequency response of the rectangular Waveguide Mesh has been motivated;
- a parametrization of 1-st order Digital Waveguide Filters has been proposed in the case when those filters are tuned to simulate real surfaces;
- a "scattered" formulation of the Waveguide Mesh boundaries has been proposed to account for multiple surface reflections over a single boundary point;
- a modified, delay-free triangular Waveguide Mesh has been designed that minimizes numerical dispersion, along with a general method that allows to compute delay-free (i.e., implicitly computable) linear filter networks without rearranging the network topology;
- a set of morphing 3-D resonant cavities has been figured out to study the perceptually significant aspects of their spectra, in order to understand whether humans *hear* 3-D shapes in between the sphere and the cube;
- a physics-based virtual listening environment has been defined for the synthesis of reliable distance cues. This model must be considered as a main result in this work, gathering most of the ideas previously developed;
- this virtual environment has been tested with promising results in experiments simulating everyday listening conditions, using both headphone and loudspeaker conventional equipment, and in the mock-up of a navigation aid device provided with an auditory-only output interface;
- a novel parametric equalization filter structure has been developed, requiring minimal parametric reconfiguration in front of the inversion of its equalization function, this advantage costing few additional computations during each time cycle;
- as an example of physics-based sound synthesis, *crumpling* has been used for synthesizing sounds resulting from the superposition of "atomic" events. By means of that approach we were able in particular to sonify crushing cans and footsteps, gaining a direct control over specific objective parameters such as size, crushing force and material stiffness.

We envision that physics-based modeling will play a role in the auditory display of the future, as soon as reliable and versatile virtual sounding and spatialization

objects will become available to the HCI community. Using those objects interface designers will be able to specify the sonic events in the scene through direct manipulation of parameters, this manipulation resulting into specific object behaviors and event dynamics. Moreover, designers will be enabled to choose where to locate those objects and events in the scene via the selection of proper spatial attributes describing the listening environment.

If this scenario becomes a reality, then the author will feel rewarded of the effort undertaken for writing this work, whatever its contribution.

Part II

Articles

List of Articles

- *Appendix A*
  Online Correction of Dispersion Error in 2D Waveguide Meshes.
  Federico Fontana and Davide Rocchesso
  Proc. International Computer Music Conference, pages 78–81, Berlin, Germany, August 2000.
- *Appendix B*
  Using the waveguide mesh in modelling 3D resonators.
  Federico Fontana, Davide Rocchesso and Enzo Apollonio
  Proc. Conference on Digital Audio Effects (DAFX-00), pages 229–232, Verona, Italy, December 2000.
- *Appendix C*
  Signal-Theoretic Characterization of Waveguide Mesh Geometries for Models of Two–Dimensional Wave Propagation in Elastic Media.
  Federico Fontana and Davide Rocchesso
  IEEE Trans. Speech and Audio Processing, 9(2):152–161, February 2001.
- *Appendix D*
  A Modified Rectangular Waveguide Mesh Structure with Interpolated Input and Output Points.
  Federico Fontana, Lauri Savioja and Vesa Välimäki
  Proc. International Computer Music Conference, pages 87–90, La Habana, Cuba, September 2001.
- *Appendix E*
  Acoustic Cues from Shapes between Spheres and Cubes.
  Federico Fontana, Davide Rocchesso and Enzo Apollonio
  Proc. International Computer Music Conference, pages 278–281, La Habana, Cuba, September 2001.
- *Appendix F*
  Recognition of ellipsoids from acoustic cues.
  Federico Fontana, Laura Ottaviani, Matthias Rath and Davide Rocchesso
  Proc. Conference on Digital Audio Effects (DAFX-01), pages 160–164, Limerick, Ireland, December 2001.
- *Appendix G*
  A Structural Approach to Distance Rendering in Personal Auditory Displays.

Federico Fontana, Davide Rocchesso and Laura Ottaviani
Proc. IEEE International Conference on Multimodal Interfaces (ICMI'02), pages 33–38, Pittsburgh, PA, October 2002.

- *Appendix H*
  A Digital Bandpass/Bandstop Complementary Equalization Filter with Independent Tuning Characteristics.
  Federico Fontana and Matti Karjalainen
  IEEE Signal Processing Letters.
  - IN PRESS -
- *Appendix I*
  Characterization, modelling and equalization of headphones.
  Federico Fontana and Mark Kahrs
  Journal of the Virtual Reality Society.
  - SUBMITTED FOR REVIEW -
- *Appendix J*
  Computation of Linear Filter Networks Containing Delay-Free Loops, with an Application to the Waveguide Mesh.
  Federico Fontana
  IEEE Trans. Speech and Audio Processing.
  - ACCEPTED FOR PUBLICATION -
- *Appendix K*
  Acoustic distance for scene representation.
  Federico Fontana, Davide Rocchesso and Laura Ottaviani
  IEEE Computer Graphics & Applications.
  - SUBMITTED FOR REVIEW -

A

Online Correction of Dispersion Error in 2D Waveguide Meshes

Federico Fontana and Davide Rocchesso
Proc. International Computer Music Conference, pages 78–81, Berlin, Germany, August 2000.

An elastic ideal 2D propagation medium, i.e., a membrane, can be simulated by models discretizing the wave equation on the time–space grid (finite–difference methods), or locally discretizing the solution of the wave equation (waveguide meshes). The two approaches provide equivalent computational structures, and introduce numerical dispersion that induces a misalignment of the modes from their theoretical positions. Prior literature shows that dispersion can be arbitrarily reduced by oversizing and oversampling the mesh, or by adopting offline warping techniques. In this paper we propose to reduce numerical dispersion by embedding warping elements, i.e., properly tuned allpass filters, in the structure. The resulting model exhibits a significant reduction in dispersion, and requires less computational resources than a regular mesh structure having comparable accuracy.

A.1 Introduction

Membranes are the crucial component of most percussion instruments. Their response to an excitation, and their interaction with the rest of the musical instrument and with the environment, strongly affect the sound quality of a percussion. Physical modeling of membranes has drawn the attention of the computer music community when a new model based on the Digital Waveguide was designed, called 2-D Digital Waveguide Mesh [157]. The model was proved to provide a computational structure equivalent to a Finite Difference Scherme (FDS). In particular, it was shown that the numerical artifacts introduced by the model cause a phenomenon called dispersion. This means that, even in a flexible medium, different spatial frequency components travel at different speeds, and this speed is direction- and frequency-dependent [152, 157].

Different mesh geometries have been studied: each of them have an equivalent FDS, and exhibits its peculiar dispersion error function [53]. The triangular geom-

etry exhibits two valuable properties: the dispersion error is, with good approximation, independent from the direction of propagation of the spatial components; at the same time, the Triangular Waveguide Mesh (TWM) defines, from a signal-theoretic point of view, the most efficient sampling scheme among the geometries that can be derived from mesh models used in practice [53,138]. The independency from direction has been successfully exploited [138] to warp the signals produced by the model, using offline filtering techniques [70]. In this paper we work on a similar idea, but warping is performed online by cascading each unit delay in the TWM with a first–order allpass filter. By properly tuning the filter parameter, we will prove that a considerable reduction of the dispersion error can be achieved in a wide range around dc.

This result is then compared with the performance of a TWM, oversized in order to reduce dispersion in the first modes. It will be shown that the warped mesh is less expensive in terms of memory and computational requirements. This evidence holds both for the straight waveguide and the FDS implementations. Our conclusion is that the most efficient, low–dispersion computational scheme for membrane modeling is a triangular FDS where the unit delays are cascaded with properly tuned allpass filters.

Having an efficient and accurate membrane model is a key step toward the construction of affordable, tunable, and realistic models of complete percussion instruments. In particular, the coupling between air and membrane [51], and the interface with resonating structures are fundamental components that should be added to the membrane model in order to achieve better realism [2].

A.2 Online Warping

For a wave traveling along the waveguide mesh, the numerical dispersion error is a function of the two spatial frequency components. In the TWM, this function is symmetric around the origin of the spatial frequency axes, with good approximation. Consequently, it makes sense to plot the dispersion factor as a single-variable function of spatial frequency, moving from the center of the surface to the absolute band edge along one of the three directions defined by the waveguide orientations[1]. Assuming the waveguides to have unit length, the spatial band edge results to be equal to $2\pi/\sqrt{3}$ [rad/spatial sample] [53]. A plot of the dispersion factor versus temporal frequency is then calculated recalling the nominal propagation speed factor, equal to $1/\sqrt{2}$ [spatial sample/time sample], affecting any finite difference model [152], that fixes the edge of the temporal frequency at the value $\sqrt{2}\pi/\sqrt{3}$ [rad/sample]. Figure A.1 depicts a plot of the dispersion factor.

This analysis is confirmed by simulations conducted over a TWM modeling a square membrane of size 24×24 waveguide sections, clamped at the four edges, excited at the central junction by an impulse. In fact, the impulse response taken at the central junction shows that the positions of its modes match well with the theoretical frequencies of the odd modes of the membrane, each one of them being shifted by its own dispersion and by the nominal propagation speed factor.

[1] In [138], a function averaging the surface magnitude around the origin is constructed, resulting in a slight difference respect to the curve adopted here.

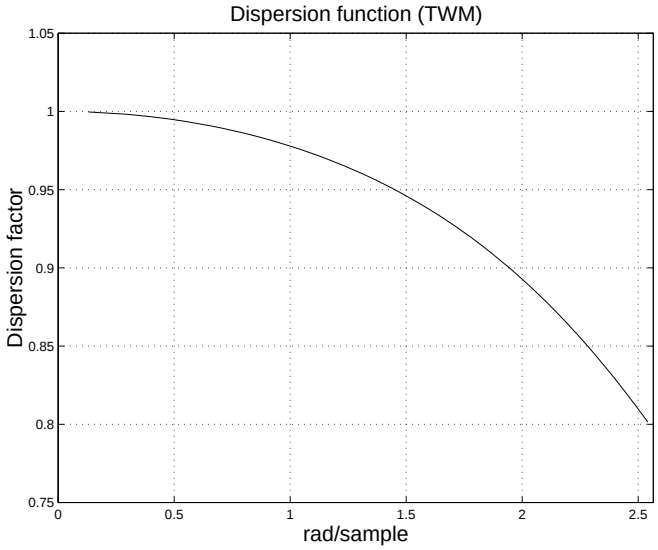

Fig. A.1. Plot of the dispersion error versus temporal frequency magnitude

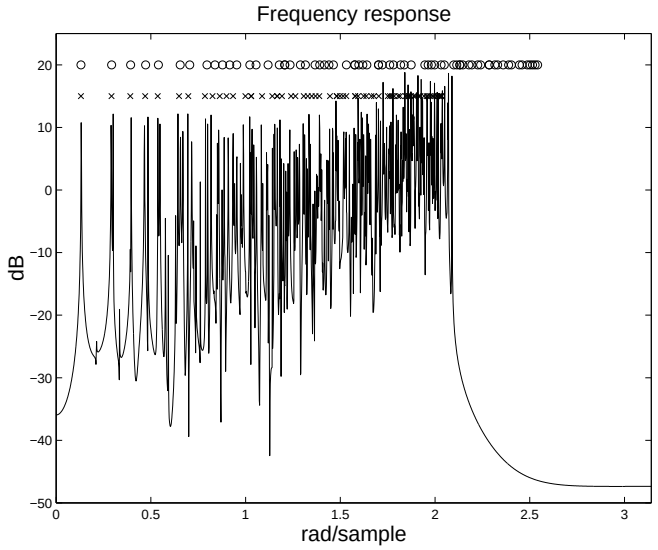

Fig. A.2. Frequency response taken at the center of a TWM (size 24×24) excited by an impulse at the same point. Theoretical positions of the odd modes resonating in a mebrane below $\sqrt{2}\pi/\sqrt{3}$ [rad/sample], weighted by the nominal propagation speed factor ($\circ$). Positions of the same modes affected by dispersion ($\times$).

The results are depicted in Figure A.2, where the frequency response of the model is plotted together with ($\circ$) the theoretical positions (compressed by the nominal propagation speed factor) of the modes resonating in the membrane below $\sqrt{2}\pi/\sqrt{3}$ [rad/sample], and ($\times$) the real positions of the same modes, affected by

dispersion. Overall, dispersion introduces a modal compression that increases with frequency. The careful reader will note a slight difference between the calculated frequency cut and the bandwidth of the signal coming from the simulation. This difference is probably due to the simplifying assumption of considering the dispersion function as direction independent. Moreover, some modes show up as twin peaks. This may be due to the actual irregular shape of the resonator model, caused by the impossibility to design a perfectly square geometry using a TWM model. In order to conduct a controlled analytical study we avoided using interpolation along the edge, even though this is recommended in practical implementations [2].

Let $H(z)$ be the transfer function of a TWM, regardless of the excitation (input) and acquisition (output) positions. The transfer function can be handled by conformal mapping, a method consisting in the application of a particular map T to the z-domain, in order to obtain a new, warped domain $\tilde{z} = T(z)$ [70, 101]. The frequency response $H(e^{j\tilde{\omega}})$, calculated from $H(\tilde{z})|_{\tilde{z}=e^{j\tilde{\omega}}}$, changes according with the map.

Practical implementations of transfer functions obtained by conformal mapping are often affected by non computable loops, that can sometimes be resolved [69]. In a TWM, delay free loops appear whenever the map does not allow to extract an explicit unit delay. However, if we change the number of unit delays in each waveguide section of a waveguide mesh, we only change the number of Fourier images of the frequency response[2], by simply compressing each single image. Now, imagine a map that translates each unit delay into the cascade of a first–order allpass filter $A(z)$ and a unit delay, $\tilde{z}^{-1} = z^{-1}A(z)$. If the allpass filter has a negative coefficient, the phase delay introduced by the filter ranges from one sample (in high frequency) to a certain value larger than one (in low frequency). Therefore, it is reasonable to expect an extra image (due to doubling the unit elements for high frequencies), and a degree of compression that decreases with increasing frequency. This is exactly the kind of behavior that is desired in order to counterbalance the effects of numerical dispersion. In order to get rid of extra frequency components it is sufficient to lowpass filter at the desired cutoff frequency, and to restore the correct positions of low-frequency partials it is sufficient to increase the temporal sampling rate.

The frequency domain is warped according with the formula

$$\tilde{\omega} = \arctan\left[\frac{2\sin\omega\{\alpha + \cos\omega\}}{1 + \alpha^2 + 2\alpha\cos\omega - 2\sin^2\omega}\right], \tag{A.1}$$

where α is the parameter of the allpass filter A(z):

$$A(z) = \frac{\alpha + z^{-1}}{1 + \alpha z^{-1}}. \tag{A.2}$$

Figure A.3 shows the mapping functions calculated for some negative values of the parameter of the allpass. The more negative is α, the more warped are the modes, especially in the low frequency range. This result has an intuitive interpretation if we consider the phase delay of the allpass: the more negative is α, the more delayed

[2] This occurs whenever a map $\tilde{z} = z^M$ is applied to a discrete-time linear filter.

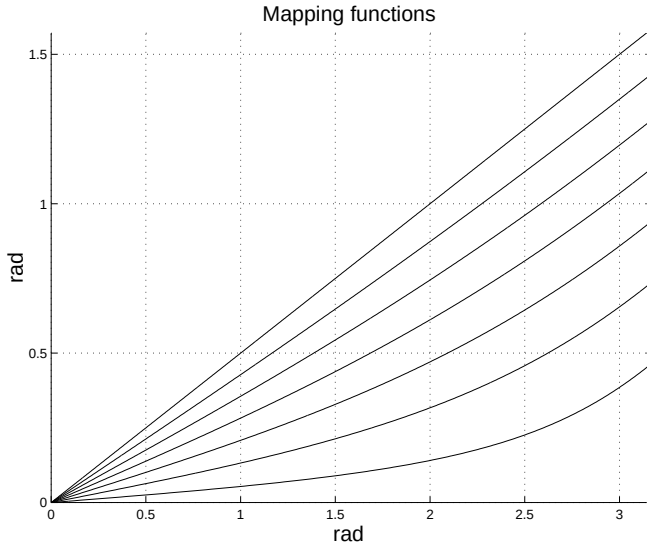

Fig. A.3. Mapping functions $\tilde{z}^{-1} = z^{-1}A(z)$ for equally-spaced values of the parameter α of the allpass filter $A(z)$. Top line: $\alpha = 0$. Bottom line: $\alpha = -0.9$.

by the allpass filter are the lower frequencies traveling into the mesh respect to the higher ones.

Choosing $\alpha = -0.45$ (curve in the middle), the warping so introduced limits the modal dispersion to very low values. Figure A.4 shows the frequency response of the same TWM simulated before, after warping. Using the same notation of figure A.2, the response is compared with the ideal positions of the modes, rescaled to align the fundamentals ($\circ$), and with the real positions of the same modes, affected by the residual dispersion ($\times$). The improvement in terms of precision in the alignment of the modes is evident by comparison of crosses and circles in figures A.2 and A.4.

A.3 Computational Performance

Figure A.5 shows a plot of the dispersion factor after warping. Dispersion is below 2% in a range equal to 75% of the whole band, then it climbs to the maximum.

From a perceptual viewpoint, it is not clear how much tolerance we might admit in the frequency positions of high order partials of a drum. Frequency deviation thresholds should be derived from subjective experimentation, as it was done for piano sounds [129]. Certainly, FigureA.2 shows an unnatural compression of modes that results in a decrease in brightness. Moreover, the frequency distribution of resonances brings us some information about the shape of the resonating object, for instance making it possible to discriminate a circular from a square drum. The warped TWM preserves the correct distribution of modes quite well up to 75% of the frequency band.

By comparison of Figures A.1 and A.5, we can note that a similar precision is achieved by a TWM if the waveguides are reduced to one third of the original

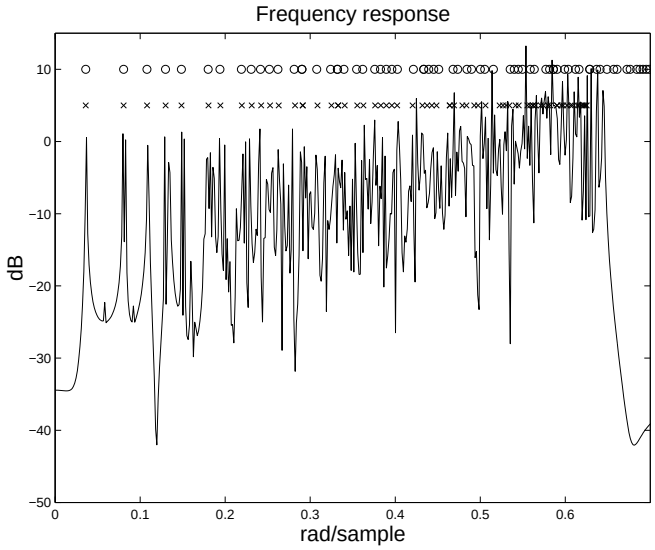

Fig. A.4. Impulse response taken at the center of a warped TWM (size 24×24) excited by an impulse at the same point. Theoretical positions of the odd modes resonating in a mebrane, rescaled to align the fundamentals ($\circ$). Positions of the same modes affected by residual dispersion ($\times$).

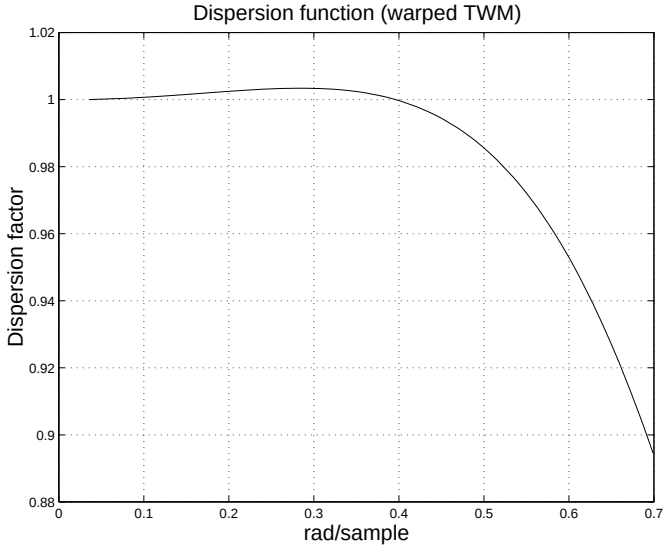

Fig. A.5. Plot of the dispersion error versus temporal frequency magnitude in the warped TWM.

length (the mesh is nine times denser). Then, the fundamentals can be aligned in the two models by multiplying times 1.75 the sampling rate of the warped TWM. Finally, both the output signals must be lowpass filtered.

	Sums	Mult	Memory
TWM	99	9	54
WTWM	40.25	22.75	22.75
FDS	54	9	18
WFDS	17.5	8.75	7

Table A.1. Performance of the TWM (FDS) vs. warped version ('W') in terms of sums, multiplies and memory locations. Both the TWM (FDS) and its warped version allow the same dispersion tolerance.

From these considerations, a comparison of the TWM versus its warped version in terms of needed sums, multiplies and memory locations, based on dispersion tolerance, can be summarized in Table A.1, where the warped models are labeled with the prefix 'W', and the allpass filter is supposed to be implemented in canonical (2 multiplies, 1 delay) form. Both the straight mesh and FDS implementations are considered. The number of multiplies can be further reduced at the expense of more memory by using one-multiply allpass filter structures.

A.4 Conclusion

A new technique to reduce modal dispersion in a wide frequency range in TWM and triangular FDS models of 2D resonators has been presented. This technique is based on first–order allpass filters embedded in the mesh, and it requires an increase in temporal sampling rate accompanied by lowpass filtering of the output signal. The resulting warped TWM is shown to be less expensive in terms of computing resources and memory consumption than oversizing a TWM or FDS model. The coefficient of the embedded allpass filters is also a parameter that can be controlled to introduce tension modulation or other more exotic effects.

B

Using the waveguide mesh in modelling 3D resonators

Federico Fontana, Davide Rocchesso and Enzo Apollonio
Proc. Conference on Digital Audio Effects (DAFX-00), pages 229–232, Verona, Italy, December 2000.

Most of the results found by several researchers, during these years, in physical modelling of two dimensional (2D) resonators by means of waveguide meshes, extend without too much difficulty to the three dimensional (3D) case. Important parameters such as the dispersion error, the spatial bandwidth, and the sampling efficiency, which characterize the behavior and the performance of a waveguide mesh, can be reformulated in the 3D case, giving the possibility to design mesh geometries supported by a consistent theory.

A comparison between different geometries can be carried out in a theoretical context. Here, we emphasize the use of the waveguide meshes as efficient tools for the analysis of resonances in 3D resonators of various shapes. For this purpose, several mesh geometries have been implemented into an application running on a PC, provided with a graphical interface that allows an easy input of the parameters and a simple observation of the consequent system evolution and the output data. This application is especially expected to give information on the modes resonating in generic 3D shapes, where a theoretical prediction of the modal frequencies is hard to do.

B.1 Introduction

Multidimensional resonators can be found in all musical instruments and in almost all listening contexts. Hence, the aim to model them and to simulate their behavior by means of stable, versatile and easy-to-handle numerical methods is strongly felt. Recently, Waveguide Meshes (WM) have been proposed and carefully studied by several researches, as structures especially devoted to model wave propagation along an ideal, multidimensional medium [50, 158].

Waveguide meshes are built by interconnecting Digital Waveguides [148] according to several topologies, and provide computational structures equivalent to a Finite Difference Scheme (FDS) [51, 157]. Hence, most of the FDS theory can be

recovered from the theoretical analysis of the WM: in particular, it can be shown that these structures introduce numerical artifacts that can be interpreted in terms of dispersion error. This means that, even in the simulation of a non-dispersive medium, different spatial frequency components travel at different speeds, and this speed is direction- and frequency-dependent. The dispersion functions vary with the topology of the mesh, but in any case they cause a misalignment of the resonant modes from their theoretical positions [152].

At the same time, the waveguide approach results to be more intuitive and meaningful in several questions related with these models, such as the design of the boundaries and the choice of the mesh topology. Using this approach, it has been recently understood that the dispersion error, introduced by discretizing in time and space the propagation medium, can be reduced in certain topologies using interpolation, and off line or online warping techniques [52, 138]. These techniques increase the cost of the resulting model less than a reduction of the dispersion obtained by mere oversizing of the mesh.

Definitely, research in WMs is leading to interesting models of resonators, even if much still needs to be done. For instance, there is lack of simulations producing natural sounds, although the WM should couple without too much difficulty with realistic damping elements and other resonating structures.

With this goal in mind, we have started to implement various WM geometries into an application running on a PC. In particular, the application embeds two new 3D geometries still unseen in the literature, for which a brief theoretical discussion is given in the next section. The application accepts parameters as mesh geometry, junctions' density, shape of the resonator, type of excitation and positions of the listening points. Together with them, more complex parameters such as the online warping factor, here obtained by cascading the digital waveguides with properly tuned all-pass filters, allow to control dispersion and/or to shift dynamically the modes. The application includes a visual interface that makes the software a useful tool for the analysis of ideal resonators and mesh structures, and a versatile building block for future developments.

B.2 3D schemes

In the 2D case, the geometries that can be chosen are not so many. Among the available ones, the literature suggests to select the triangular or the de-interpolated WM, which have the most regular behavior of the dispersion, or the square geometry, which is a good trade-off between computational cost and simplicity of implementation of its 4-port junctions [52, 138].

Once the third dimension is added, the collection of available geometries becomes richer: it includes the orthogonal WM, already exploited in simulation of room acoustics [134], the tetrahedral WM [160], and other, sometimes complex, topologies. In order to make a good choice in this panorama, it should be proved, as done in the 2D case [53], that

- geometries in which the junctions have more than one orientation (i.e.: the tetrahedral) in general do not translate into efficient models, due to the gaps present in the underlying sampling scheme;

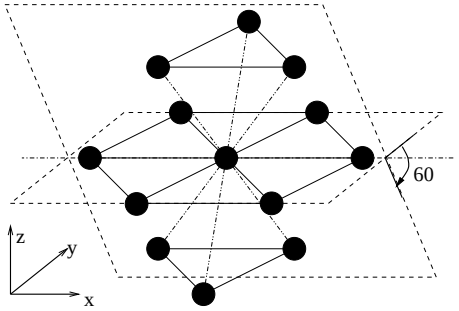

Fig. B.1. The 3DTWM.

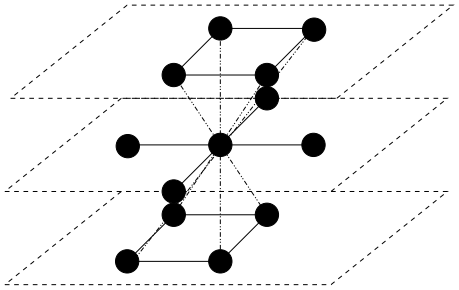

Fig. B.2. Superposition of interlaced square WMs.

- meshes where the signals, traveling from one (input) junction to another (output) junction, run through a number of waveguides which is always even or always odd, define input-output transfer functions in the variable z^{-2}. Hence, their frequency response mirrors at half the Nyquist frequency;
- geometries resulting in particular distributions of the junctions in the 3D space, map themselves into non-orthogonal, often efficient sampling schemes;
- WMs exhibiting a uniform characteristic of the dispersion error, i.e. a good degree of symmetry of the dispersion function around the origin, allow a reduction of this error by means of frequency warping techniques.

In particular, we have chosen two geometries satisfying these assumptions. The former, which we are calling 3D triangular WM (3DTWM), is presented in Figure B.1: it is obtained by superposing four triangular WMs, through rotations of 60° of one triangular WM along the three axes parallel to the waveguide directions (in Figure B.1, one plan obtained by the rotation along one of these axes is depicted). The latter is defined by the superposition along the third dimension of planes containing square WMs, in such a way that each couple of adjacent planes is shifted of half the waveguide length in both waveguide direction (see Figure B.2), so creating interlaced adjacent square WMs, separated by a distance tuned to equate the length of all the waveguides composing the 3D structure.

Both the geometries are made of junctions, each one having only one orientation. Both schemes allow walks of the signal through even or odd numbers of waveguides. Both of them lie on efficient, non-orthogonal sampling schemes. Finally, both of them — especially the 3DTWM — have a very uniform dispersion

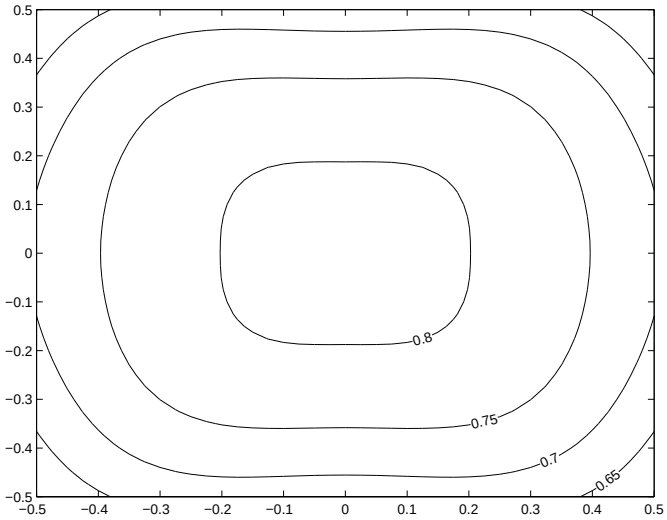

Fig. B.3. Projection of the dispersion function over the plan (x, z) of the normalized frequency domain in the 3DTWM.

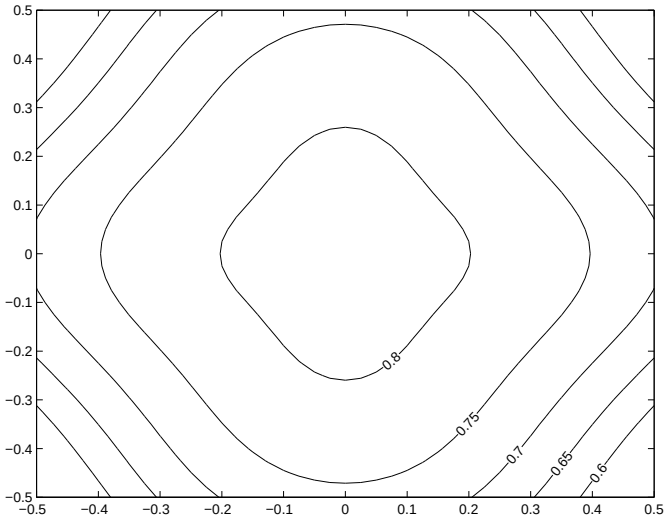

Fig. B.4. Projection of the dispersion function over the plan (x, z) of the normalized frequency domain in the superposition of interlaced square WMs.

characteristic, as shown in Figures B.3 and B.4, where projections of the dispersion function over the plan (x, z) of the normalized frequency domain are plotted.

In conclusion, the two geometries can be adopted in practice as efficient structures to model 3D resonators.

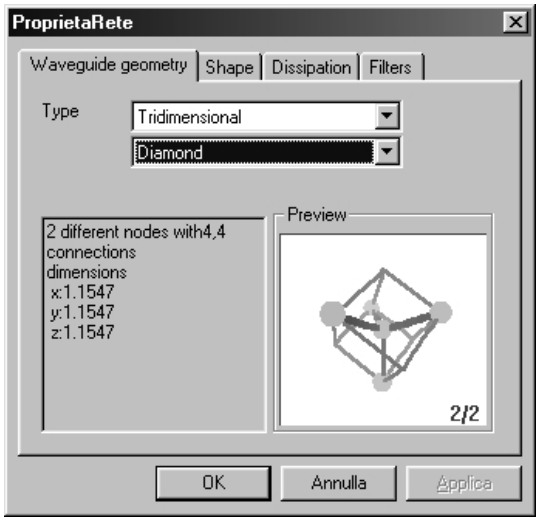

Fig. B.5. The input window in the application.

B.3 Implementation

Several geometries have been implemented in an application — still unnamed — written in C++ language. The scattering junctions, together with their waveguide neighbors, have been designed as objects, in such a way that a mesh is generated using a constructor, which builds up the structure according to the dimensions and the shape of the resonator, and to the mesh geometry. For this reason, the resonator turns out to be a composition of atomic volumes, where each volume is filled with one object. In this way it is possible to compose resonators with given shape and dimensions, using the desired geometry of the WM (Figure B.5).

Further parameters may be selected. In particular, a coefficient of attenuation, which controls the decay of the signals, can be set. Meanwhile, a warping coefficient, controlling the parameter of first-order all pass filters embedded in the meshes, can be tuned according with the characteristics of the resonator. Finally, one or more input signals can be injected in any junction; in particular, when a 2D circular resonator of radius R and tension T is modeled, the input function of displacement d can be set in the form

$$d(r) = \begin{cases} \dfrac{F}{2\pi T}\left[\ln\dfrac{R}{s} + \dfrac{1}{2s^2}(s^2 - r^2)\right] & , 0 \le r < s \\ \dfrac{F}{2\pi T}\ln\dfrac{R}{s} & , s \le r < R \end{cases},$$

corresponding to the initial condition for the displacement in a point r units far from the center, when a uniform force F is applied to an area of radius s centered over the membrane [102].

The system evolution can be monitored step by step by direct observation of the signal in correspondence of the junctions, in several ways. In particular, the

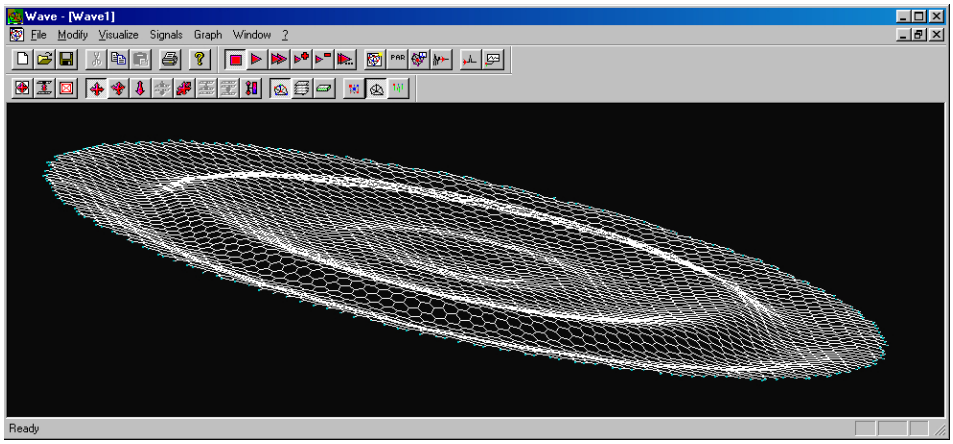

Fig. B.6. The application. The visual interface allows to monitor the system evolution.

signal traveling inside a 3D resonator can be observed by the colors assumed by the junctions, or by visualization of "slices" of the resonator, each one being one of the superimposed plans composing the solid shape. Otherwise, the application can be made running indefinitely.

Similarly, one or more output junctions can be chosen in the mesh. The output signals or, equivalently, their spectra, can be monitored during the system evolution, and saved as raw files that can be easily transformed into new data structures for feeding more complex analysis tools as Matlab^{TM}.

The application, still at its first steps, has already allowed to do some interesting observations about the evolution of the wave signals traveling along the mesh. In particular, it has confirmed the high degree of redundancy present in the signal, when low-efficiency geometries are selected. In the 3D case, this implies a large amount of operations that can be avoided adopting more efficient schemes.

Soon, a functionality for the creation of general parametric shapes will be embedded in the application. Hybrid resonators will be obtained by generating mesh structures according with the equations that express possible transitions between shapes, whose resonance modes are well known in physical acoustics (like, for example, a cube or a sphere). By this improvement in the application, we expect to study the sounds coming out from hybrid 3D resonators, inspired in this by studies on the perception of spatial relationships conducted in the field of visual perception and image synthesis [164] and, at a starting level, in the field of psychophysical acoustics [126].

By exploiting the object nature of the software, further structures could be embedded in the code, to implement other typical features which characterize the resonators, like wall absorption or wave diffusion, and to couple two or more different resonators. This goal needs further theoretical investigations about the best way to integrate these features into the waveguide paradigm, and could be hopefully achieved by sharing the code between all the research communities interested in the analysis and the application of the WM.

B.4 Summary

Waveguide meshes are coming to a point of maturity. The underlying theory allows now to determine the best geometry in the 2D case, respect to requirements of precision and cost of the model. The same theory should extend with not much effort to the third dimension, so helping in determining the most efficient and useful geometries devoted to study the resonances produced by a 3D resonator.

The implementation of several geometries into an application running on a common PC seems to confirm this assumption. Moreover, the same application shows interesting aspects in the system evolution of the waveguide models of 3D resonators, such that it can be considered as a first, general building block in the realization of a more complete simulator of waveguide models.

The object structure of the code makes this application a versatile software, where further features will be hopefully added, once the theory of waveguide meshes will be developed toward the solution of aspects related with the modeling of real resonators, holding the condition that a community of researchers will be interested in the growth of the applications for waveguide mesh simulation.

C

Signal-Theoretic Characterization of Waveguide Mesh Geometries for Models of Two–Dimensional Wave Propagation in Elastic Media

Federico Fontana and Davide Rocchesso
IEEE Trans. Speech and Audio Processing, 9(2):152–161, February 2001.

Waveguide Meshes are efficient and versatile models of wave propagation along a multidimensional ideal medium. The choice of the mesh geometry affects both the computational cost and the accuracy of simulations. In this paper, we focus on 2D geometries and use multidimensional sampling theory to compare the square, triangular, and hexagonal meshes in terms of sampling efficiency and dispersion error under conditions of critical sampling. The analysis shows that the triangular geometry exhibits the most desirable tradeoff between accuracy and computational cost.

C.1 Introduction

Among the techniques for modeling wave propagation in multidimensional media, the *Digital Waveguide Meshes* have recently been established [50, 51, 135, 157, 158, 160, 161] as intuitive and efficient formulations of finite difference methods [152].

A Waveguide Mesh (WM) is a discrete–time computational structure that is constructed by tiling a multidimensional medium into regular elements, each giving a local description of wave propagation phenomena. This local description is lumped into a waveguide junction [147, 158], which is lossless by construction. Therefore, waveguide meshes are free of numerical losses, even though lumped passive elements can be explicitely inserted to simulate physical losses. However, wavefronts propagating along a multidimensional WM are affected by dispersion error, i.e. different frequencies experience different propagation velocities. Numerical dispersion cannot be completely eliminated, but it can be arbitrarily reduced increasing the density of the elements, and minimized choosing the "least dispersive geometry". Moreover, interpolation schemes or off-line warping techniques [137] can be applied to attenuate the effects of numerical dispersion.

In this paper we investigate how the density of waveguide junctions and the sampling frequency affect the signal coming out from the model. As the analysis depends on the geometry of the WM, we focus on 2D media, finding properties

for the square, triangular and hexagonal WM (named respectively SWM, TWM and HWM in the following). Such properties allow to calculate the bandwidth of a signal produced by a WM working at a given sampling frequency, once its geometry and the density of its junctions have been determined.

The paper is structured as follows. Section C.2 provides some background material on waveguide meshes (and their interpretation as finite difference schemes) and multidimensional sampling lattices. Section C.3 illustrates the spatial sampling efficiency of the three waveguide mesh geometries for signals having circular spatial band shape. In Section C.4 we explain how the critical spatial sampling affects the choice of the temporal sampling frequency in non-aliasing conditions. In Section C.5, the computational performances of the three geometries are compared under critical sampling conditions.

C.2 Background

C.2.1 Digital Waveguides and Waveguide Meshes

An ideal one-dimensional physical waveguide can be modeled, in discrete time, by means of a couple of parallel delay lines where two wave signals, s_+ and s_-, travel in opposite directions. Such a structure, based on spatial sampling (with interval D) and time sampling (with interval T), is called a digital waveguide [148,149]. If wave propagation in the physical medium is lossless and non-dispersive with speed $c = D/T$, no error is introduced by the discrete-time simulation as long as s_+ and s_- are band limited to a band $B = (2T)^{-1}$. In this case, the signal $s(x,t)$ along the physical waveguide can be reconstructed with no aliasing error from samples of the wave signals:

$$s(mD, nT) = s_+(mD, nT) + s_-(mD, nT) \; . \tag{C.1}$$

N digital waveguide terminations can be connected by means of a lossless scattering junction [147,158]. Preservation of the total energy in the form of Kirchhoff's node equations leads to the scattering equation

$$s_{i-} = \frac{2}{N} \sum_{k=1}^{N} s_{k+} - s_{i+} \quad i = 1, \ldots, N \; , \tag{C.2}$$

which allows for calculating the outgoing wave signal to the i-th waveguide branch from the N incoming wave signals $s_{1+}, \ldots, s_{N+}$, under the assumption of equal wave impedance at the junction for all the waveguides.

A WM, as proposed by Van Duyne and Smith in 1993 [157,158], is obtained by connecting unit-length digital waveguide branches by means of lossless scattering junctions. For the simulation of uniform and isotropic multidimensional media, a few kinds of geometries, all corresponding to tiling the multidimensional space into regular elements, have been proposed: square [158], triangular [50,161], hexagonal [161] for 2D media such as membranes; rectilinear [135] and tetrahedral [160,161] for 3D media. The WMs which are considered in this paper (SWM, TWM and HWM) are depicted in Figure C.1. Among the proposed geometries, the HWM

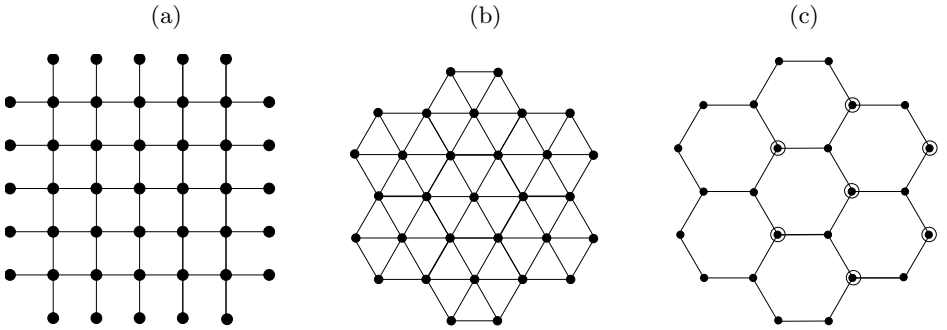

Fig. C.1. The SWM (a), the TWM (b) and and the HWM (c). $-$,/,\ and | are digital waveguides, $\bullet$ are lossless scattering junctions. In (c), seven lossless scattering junctions separated by two digital waveguide branches are marked with $\odot$.

is peculiar because its 3–port lossless scattering junctions exhibit two different orientations, and it can be interpreted as two interlaced TWMs (see the junctions marked with $\odot$ in Figure C.1 (c).

WMs introduce a basic relation, between signal $s(\mathbf{x}_j, nT)$, taken from a junction located at position $\mathbf{x}_j$ in space, and signals $s(\mathbf{x}_j + \mathbf{D}_k, nT)$, $k = 1, \ldots, N$, taken from the N adjacent junctions connected to it, D meters far from position $\mathbf{x}_j$:

$$s(\mathbf{x}_j, nT + T) + s(\mathbf{x}_j, nT - T) = \frac{2}{N} \sum_{k=1}^{N} s(\mathbf{x}_j + \mathbf{D}_k, nT) \tag{C.3}$$

which is obtained from (C.2).

Equation (C.3) indicates that each WM behaves like a finite difference scheme [152], the difference being that the former has a state lumped in the digital waveguides, and the latter has a state lumped in the junctions. For the purpose of the analysis that follows, when the lossless scattering junctions have more than one orientation, as in the HWM, a difference equation such as (C.3) should be written as many times as there are orientations, each one using a proper set of vectors $\mathbf{D}_1, \ldots, \mathbf{D}_N$ [161].

Following the lines of von Neumann stability analysis [152], Equation (C.3) can be Fourier transformed with spatial variables x and y, resulting in

$$S(\xi_x, \xi_y, nT + \alpha_{\mathrm{g}} T) + S(\xi_x, \xi_y, nT - \alpha_{\mathrm{g}} T) = b_{\mathrm{g}} S(\xi_x, \xi_y, nT) \tag{C.4}$$

where ξ_x and ξ_y are spatial frequencies, α_{g} takes the value 2 for the HWM and 1 for the other meshes, and b_{g} is a geometric factor equal to

$$
\begin{aligned}
b_{\mathrm{s}} &= \cos(2\pi D\xi_x) + \cos(2\pi D\xi_y) \\
b_{\mathrm{t}} &= \frac{2}{3}\cos\left(2\pi D\xi_x\right) + \frac{2}{3}\cos\left(2\pi D\left[\frac{1}{2}\xi_x + \frac{\sqrt{3}}{2}\xi_y\right]\right) \\
&\quad + \frac{2}{3}\cos\left(2\pi D\left[\frac{1}{2}\xi_x - \frac{\sqrt{3}}{2}\xi_y\right]\right) \\
b_{\mathrm{h}} &= \frac{8}{9}\cos\left(2\pi\sqrt{3}D\xi_x\right) + \frac{8}{9}\cos\left(2\pi D\left[\frac{\sqrt{3}}{2}\xi_x + \frac{3}{2}\xi_y\right]\right) \\
&\quad + \frac{8}{9}\cos\left(2\pi D\left[\frac{\sqrt{3}}{2}\xi_x - \frac{3}{2}\xi_y\right]\right) - \frac{2}{3}
\end{aligned}
\tag{C.5}
$$

for the SWM, TWM and HWM, respectively.

Solving Equation (C.4) as a finite difference equation in the discrete–time variable, the spatial phase shift affecting a traveling signal in one time sample is found to be

$$
\Delta\varphi_{\mathrm{g}}(\xi_x, \xi_y) = -\frac{1}{\alpha_{\mathrm{g}}}\arctan\frac{\sqrt{4 - b_{\mathrm{g}}^2}}{b_{\mathrm{g}}}\,. \tag{C.6}
$$

We can compare the propagation speed of a signal traveling along a WM, versus the propagation speed of a signal traveling along an ideal membrane. If we consider membranes where relation $D = cT$ holds, meaning that signals, during a time period T, travel for a distance equal to the digital waveguide length, we can calculate the spatial phase shift of these signals, occurring during a time period:

$$
\Delta\varphi = -2\pi D\xi\,, \tag{C.7}
$$

where $\xi = \sqrt{\xi_x^2 + \xi_y^2}$. By comparing (C.6) and (C.7), we find the ratio k_{g} between the propagation speed of a signal traveling along a WM and along an ideal membrane [51, 152, 157]:

$$
k_{\mathrm{g}}(\xi_x, \xi_y) = \frac{1}{2\pi\alpha_{\mathrm{g}} D\xi}\arctan\frac{\sqrt{4 - b_{\mathrm{g}}^2}}{b_{\mathrm{g}}}\,. \tag{C.8}
$$

Since k is a non constant function of the spatial frequencies, WMs introduce a dispersion error, i.e., the signal traveling along a WM is affected by dispersion of its components. As a starting point, dispersion can be evaluated for spatial frequencies lower than the Nyquist limit[1], that is, in the frequency domain:

[1] For the purpose of this paper, the Nyquist limit is defined as half the sample rate, both in time and space. In some previous works [137, 161] the simulations were considered valid up to a quarter of the sample rate because in the square mesh the frequency response repeats itself after that limit. However, as it was pointed out in [161], this is due to the fact that all transfer functions definable at any one junction are functions of z^{-2}. This does not imply that the response to a signal having components up to half the sample rate will be aliased. However, for certain applications such as modal analysis of physical membranes, the frequency response at quarter of the sample rate is the definitive limit with the square mesh.

$$\left\{(\xi_x, \xi_y) : |\xi_x| < \frac{1}{2D}\,,\, |\xi_y| < \frac{1}{2D}\right\}\,. \tag{C.9}$$

This domain will be refined, according with the considerations to be presented in section C.3.

Figure C.2 shows plots of k_s, k_t and k_h, where D has been set to unity. It can be noticed that the propagation speed decreases for increasing spatial frequencies. In particular, k_s is maximum when $\xi_x = \xi_y$, and minimum for high values of the spatial frequencies located along the main axes, suggesting that dispersion in the SWM does not affect the diagonal components traveling along it. On the contrary, the HWM exhibits the flattest dispersion error on the region centered around dc. Finally, the TWM seems to have the most uniform behavior of the dispersion error.

The propagation speed in all the WMs has a maximum at dc:

$$k_\mathrm{g}(0) \triangleq \lim_{\xi \to 0} k_\mathrm{g}(\xi) = \frac{1}{\sqrt{2}} \quad . \tag{C.10}$$

It is worth noticing that this value corresponds with the nominal propagation speed of a signal traveling along a finite difference scheme [152].

C.2.2 Sampling Lattices

The evaluation of the dispersion error does not give a complete description of the constraints holding when an ideal membrane is modeled using WMs. In particular, a method is needed for computing the signal bandwidth a WM is able to process. The theory of sampling lattices [42, 156], that we are briefly reviewing in this section, gives the background for characterizing WMs from this viewpoint.

Let us sample a 2D continuous signal s over a domain L, subset of $\mathcal{R}^2$, so defining a discrete signal $s_L(\mathbf{x})$, $\mathbf{x} \in L$. If L can be described by means of a nonsingular matrix $\mathbf{L}$ such that each element of the domain is a linear combination of the columns of $\mathbf{L}$, the coefficients being signed integers:

$$\mathbf{x} = \mathbf{L} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}\,,\, u_1 \in \mathcal{Z}\,,\, u_2 \in \mathcal{Z}\,, \tag{C.11}$$

then L is called a *sampling lattice*, and $\mathbf{L}$ is its *basis*.

The number of samples per unit area is [42, 156]:

$$\mathcal{D}_L = \frac{1}{\det(\mathbf{L})}\,, \tag{C.12}$$

as $\mathbf{L}$ contains information about the distance between adjacent samples.

S_L, the Fourier transform of s_L, is defined over $\mathcal{R}^2$ and obtained by periodic imaging of S, Fourier transform of s. These Fourier images are centered around the elements of the lattice L^*, which is described by the basis $\mathbf{L}^{-T}$, inverse transposed of $\mathbf{L}$. Notice that the denser the sampling of s is, the sparser the image centers are. In formulas,

$$\mathcal{D}_{L^*} = \frac{1}{\det(\mathbf{L}^{-T})} = \det(\mathbf{L}) = \frac{1}{\mathcal{D}_L}\,. \tag{C.13}$$

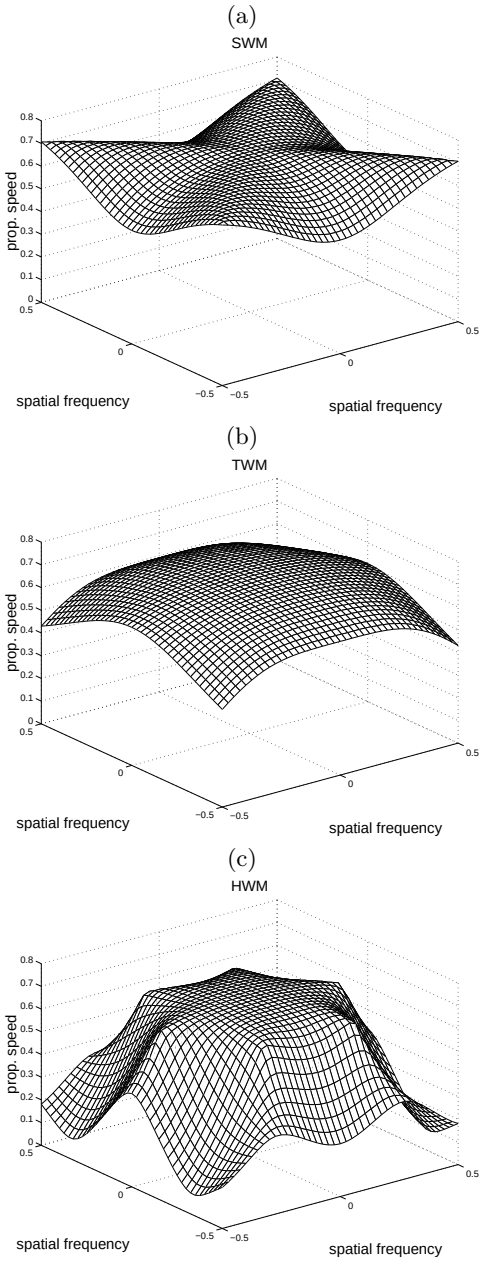

Fig. C.2. Propagation speed ratios in the SWM (a), TWM (b) and HWM (c), versus domain given by (C.9).

Let us consider an ideal, unlimited membrane traveled by a spatially band limited signal $s(x, y, t)$, and let us do a spatial sampling of the signal, at a given time. Whenever the spatial sampling defines a sampling lattice L, so that s_L is defined, we can calculate S_L using the above results. The multidimensional

sampling theorem [42] — which, in brief, tells that if the Fourier images of s do not intersect one with each other, s can be recovered from s_L with no aliasing error — indicates whether the chosen sampling scheme induces aliasing.

Equation (C.13) tells that for each choice of the sampling lattice geometry, the density $\mathcal{D}_L$ can be increased until the images do not intersect. Conversely, given s, there exists a sampling lattice capable to capture all the information needed to recover the original signal using the least density of samples. Clearly, it will exhibit the highest sampling efficiency.

C.3 Sampling efficiency of the WMs

Sampling efficiency will be calculated for signals having circular spatial band shape centered around the origin of the frequency axes[2], with radius equal to B. Even if the analysis procedure does not depend on the shape of the spatial domain, the circular band seems to include all the signals occurring in practical applications. Indeed, the procedure is independent of the system evolution — which in the WMs is controlled by Equation (C.2) — so that it applies to any model discretizing distributed systems, and where signal information can be located over a sampling lattice. This is the case, for example, in finite difference schemes.

C.3.1 WMs and Sampling Lattices

A SWM, having digital waveguides of length D_s, corresponds to a sampling scheme over the lattice $L_\mathrm{s}(D_\mathrm{s})$, described by the basis

$$\mathbf{L}_\mathrm{s}(D_\mathrm{s}) = \begin{vmatrix} D_\mathrm{s} & 0 \\ 0 & D_\mathrm{s} \end{vmatrix} . \tag{C.14}$$

A TWM, having digital waveguides of length D_t, corresponds to a sampling scheme over the lattice $L_\mathrm{t}(D_\mathrm{t})$, described by the basis

$$\mathbf{L}_\mathrm{t}(D_\mathrm{t}) = \begin{vmatrix} D_\mathrm{t} & \frac{1}{2}D_\mathrm{t} \\ 0 & \frac{\sqrt{3}}{2}D_\mathrm{t} \end{vmatrix} . \tag{C.15}$$

Notice that a triangular scheme is denser than a square scheme made with digital waveguides of the same length. This is confirmed by relation

$$\frac{\mathcal{D}_{L_\mathrm{t}(D)}}{\mathcal{D}_{L_\mathrm{s}(D)}} = \frac{\det(\mathbf{L}_\mathrm{s}(D))}{\det(\mathbf{L}_\mathrm{t}(D))} = \frac{D^2}{\frac{\sqrt{3}}{2}D^2} = \frac{2}{\sqrt{3}} . \tag{C.16}$$

The description of an HWM having digital waveguides of length D_h, again can be given in terms of sampling lattices, with some extra care. The HWM is obtained by subtracting a TWM, whose junctions lie on the sampling lattice $L_\mathrm{T}(D_\mathrm{h})$, from

[2] This class of signals encompasses the signals obtained when a membrane is excited, in one or several points, by a single shot or by a sequence of shots, using a stick with an approximately round tip.

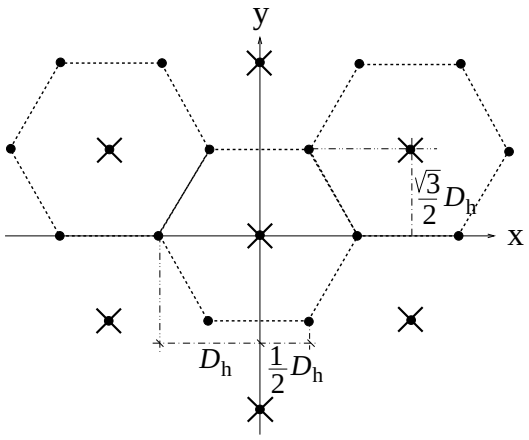

Fig. C.3. The HWM, obtained by subtraction of TWMs. $\times$ are elements belonging to $L_\mathrm{T}(D_\mathrm{h})$, $\bullet$ are elements belonging to $L_\mathrm{t}(D_\mathrm{h})$.

a denser TWM having junctions on $L_\mathrm{t}(D_\mathrm{h})$. Hence, the junctions of the HWM lie on

$$L_\mathrm{h}(D_\mathrm{h}) \stackrel{\triangle}{=} L_\mathrm{t}(D_\mathrm{h}) \setminus L_\mathrm{T}(D_\mathrm{h}) \; . \tag{C.17}$$

The basis of $L_\mathrm{T}(D_\mathrm{h})$ is

$$\mathbf{L}_\mathrm{T}(D_\mathrm{h}) = \begin{vmatrix} \frac{3}{2}D_\mathrm{h} & 0 \\ \frac{\sqrt{3}}{2}D_\mathrm{h} & \sqrt{3}D_\mathrm{h} \end{vmatrix} \; . \tag{C.18}$$

Figure C.3 shows the HWM over $L_\mathrm{h}(D_\mathrm{h})$, obtained by subtracting the sampling lattice $L_\mathrm{T}(D_\mathrm{h})$ (whose elements are depicted with $\times$) from $L_\mathrm{t}(D_\mathrm{h})$ (elements depicted with $\bullet$).

C.3.2 TWM vs SWM

The image centers of the spectra belonging to signals traveling along a SWM and a TWM are respectively described by the basis matrices of L_s^* and L_t^*:

$$\mathbf{L}_\mathrm{s}(D_\mathrm{s})^{-T} = \begin{vmatrix} 1/D_\mathrm{s} & 0 \\ 0 & 1/D_\mathrm{s} \end{vmatrix} \; , \tag{C.19}$$

and

$$\mathbf{L}_\mathrm{t}(D_\mathrm{t})^{-T} = \begin{vmatrix} \frac{1}{D_\mathrm{t}} & 0 \\ -\frac{1}{\sqrt{3}}\frac{1}{D_\mathrm{t}} & \frac{2}{\sqrt{3}}\frac{1}{D_\mathrm{t}} \end{vmatrix} \; . \tag{C.20}$$

Figure C.4 shows Fourier images for the SWM (empty circles in dashed line) and the TWM (filled circles), located around the origin of the frequency plane. It can be noticed that the square sampling scheme induces a square positioning of the images and, consequently, a square tiling of the frequency plane, as emphasized by the square in dashed line. Similarly, sampling over a TWM results in a triangular positioning of the images, thus producing an hexagonal tiling of the frequency

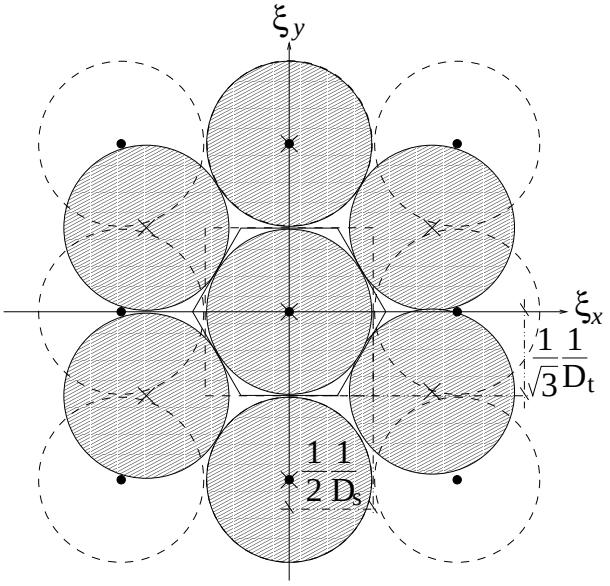

Fig. C.4. Domains of Fourier images in a SWM (empty circles in dashed line) and in a TWM (filled circles), and the correspondent tiling induced by the respective geometries (square in dashed line and hexagon in solid line). The empty and the filled circles have the same radius and touch each other without intersecting, meaning that both the SWM and the TWM critically sample the same signal.

plane, as shown by the hexagon located over the center. Such hexagonal tiling allows, as it appears quite evidently in the figure, to "pack" the images better than those coming from a square sampling.

When both a SWM and a TWM critically sample the same signal (i.e. the filled circles touch each other without intersecting, and the same thing happens for the empty circles), it is easy to derive the following relation between the digital waveguide lengths:

$$\frac{D_\mathrm{t}}{D_\mathrm{s}} = \frac{2}{\sqrt{3}} \approx 1.134 \; . \tag{C.21}$$

Under this condition we can relate the sample densities in the two geometries:

$$\frac{\mathcal{D}_{L_\mathrm{t}(D_\mathrm{t})}}{\mathcal{D}_{L_\mathrm{s}(D_\mathrm{s})}} = \frac{\mathcal{D}_{L_\mathrm{t}(\frac{2}{\sqrt{3}}D_\mathrm{s})}}{\mathcal{D}_{L_\mathrm{s}(D_\mathrm{s})}} = \frac{\det(\mathbf{L}_\mathrm{s}(D_\mathrm{s}))}{\det(\mathbf{L}_\mathrm{t}(\frac{2}{\sqrt{3}}D_\mathrm{s}))} = \frac{\sqrt{3}}{2} \; , \tag{C.22}$$

concluding that *the TWM exhibits a better sampling efficiency relative to the SWM.* In other words, a signal can be spatially sampled with a triangular geometry using 13.4% less samples per unit area.

C.3.3 TWM vs HWM

Said s_h and s_t the signals sampled over the junctions of the HWM and the TWM, respectively, and said s_T the signal sampled over the lattice $L_\mathrm{T}(D_\mathrm{h})$, we can define the zero-padded signals

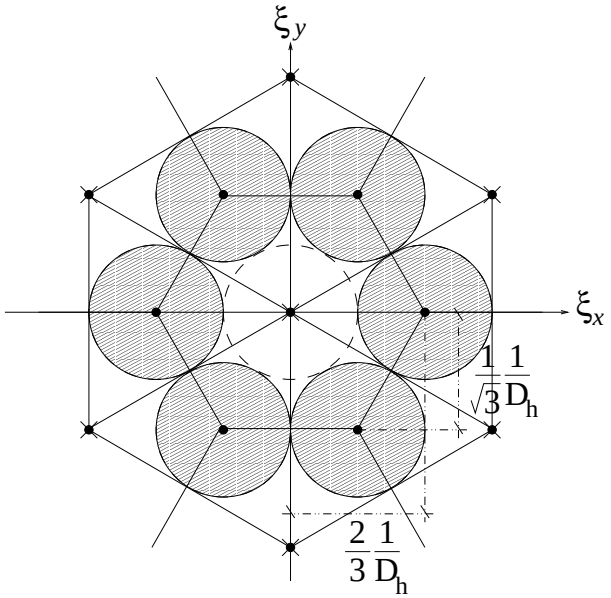

Fig. C.5. Centers ($\times$) of the Fourier images coming from a TWM defined by $L_{\mathrm{t}}(D_{\mathrm{h}})$, and centers ($\bullet$) of the Fourier images coming from a sparser TWM defined by $L_{\mathrm{T}}(D_{\mathrm{h}})$. Subtraction of $\times$ from $\bullet$ gives the centers of the Fourier images (located on the filled circles) coming from an HWM defined by $L_{\mathrm{h}}(D_{\mathrm{h}})$. The respective tiling is given by the triangles. Both the HWM and the sparser TWM critically sample the same signal.

$$\tilde{s}_{\mathrm{T}}(\mathbf{x}) = \begin{cases} 0 & , \mathbf{x} \in L_{\mathrm{h}}(D_{\mathrm{h}}) \\ s_{\mathrm{T}}(\mathbf{x}) & , \mathbf{x} \in L_{\mathrm{T}}(D_{\mathrm{h}}) \end{cases} \tag{C.23}$$

and

$$\tilde{s}_{\mathrm{h}}(\mathbf{x}) = s_{\mathrm{t}}(\mathbf{x}) - \tilde{s}_{\mathrm{T}}(\mathbf{x}) = \begin{cases} s_{\mathrm{h}}(\mathbf{x}) & , \mathbf{x} \in L_{\mathrm{h}}(D_{\mathrm{h}}) \\ 0 & , \mathbf{x} \in L_{\mathrm{T}}(D_{\mathrm{h}}) \end{cases} \tag{C.24}$$

Since we cannot define a Fourier transform over $L_{\mathrm{h}}(D_{\mathrm{h}})$, we consider $\tilde{S}_{\mathrm{h}}$, Fourier transform of $\tilde{s}_{\mathrm{h}}$ (which is defined over $L_{\mathrm{t}}(D_{\mathrm{h}})$), as a description of s_{h} in the frequency domain. $\tilde{S}_{\mathrm{h}}$, according to its definition, can be obtained by subtracting $\tilde{S}_{\mathrm{T}}$ from S_{t}:

$$\tilde{S}_{\mathrm{h}}(\xi_x, \xi_y) = S_{\mathrm{t}}(\xi_x, \xi_y) - \tilde{S}_{\mathrm{T}}(\xi_x, \xi_y) \; . \tag{C.25}$$

The result is shown in Figure C.5, where some images of S_{T} are depicted (filled circles plus circle in dashed line). The elements of ${L_{\mathrm{T}}}^*(D_{\mathrm{h}})$ are marked with $\bullet$, while the elements of ${L_{\mathrm{t}}}^*(D_{\mathrm{h}})$ are marked with $\times$.

The HWM induces an hexagonal positioning of the Fourier images (corresponding to the filled circles). In fact, their centers are elements of a set which, again, can be defined by a subtraction between two sampling lattices, ${L_{\mathrm{t}}}^*(D_{\mathrm{h}})$ and ${L_{\mathrm{T}}}^*(D_{\mathrm{h}})$, which are reciprocal of $L_{\mathrm{t}}(D_{\mathrm{h}})$ and $L_{\mathrm{T}}(D_{\mathrm{h}})$, respectively. The consequent tiling geometry is triangular, as emphasized by the triangles depicted in Figure C.5.

It can be observed that *image intersection does not occur in* $L_{\mathrm{h}}(D_{\mathrm{h}})$ *if and only if* s *is sampled without aliasing over* $L_{\mathrm{T}}(D_{\mathrm{h}})$. In other words, when S_{T} exhibits

superposition of its images, the same thing happens for the images of $\tilde{S}_{\mathrm{h}}$, and vice versa.

Moreover, it must be noticed that equation

$$L_{\mathrm{T}}(D_{\mathrm{h}}) = L_{\mathrm{t}}(\sqrt{3}D_{\mathrm{h}}) \tag{C.26}$$

holds between L_{T} and L_{t}, if a rotation is neglected. This relation, together with the considerations about image superposition made just above, allows us to say that an HWM does not perform better than a TWM, whose digital waveguides are $\sqrt{3}$ times longer.

Another interesting consideration comes out by noticing that, since

$$\frac{\mathcal{D}_{L_{\mathrm{t}}(D_{\mathrm{h}})}}{\mathcal{D}_{L_{\mathrm{T}}(D_{\mathrm{h}})}} = \frac{\det(\mathbf{L}_{\mathrm{T}}(D_{\mathrm{h}}))}{\det(\mathbf{L}_{\mathrm{t}}(D_{\mathrm{h}}))} = 3\,, \tag{C.27}$$

then the TWM defined over the lattice $L_{\mathrm{T}}(D_{\mathrm{h}})$ is 3 times as sparse as the TWM defined over $L_{\mathrm{t}}(D_{\mathrm{h}})$, thus (directly from the definition of L_{h}) 2 times as sparse as the HWM. Hence, the number of lossless scattering junctions per unit area of the HWM is twice as high as that of the TWM defined over $L_{\mathrm{T}}(D_{\mathrm{h}})$, with no benefits on the accuracy of sampling. This relates with the fact that the hexagons tiling the frequency plane (as the inner hexagon depicted in Figure C.5) contain twice the information needed to recover s correctly: the 6 partial images included in the slices inside the central hexagon can be composed into 2 Fourier images.

C.4 Signal time evolution

In this section, we discuss the consequences of critical spatial sampling on the temporal sampling frequency.

First, let us consider the sampled version of a signal traveling at speed c along an ideal membrane that has been excited by a bandlimited signal $f(x, y)$ having Fourier transform equal to $F(\xi_x, \xi_y)$. It can be shown (see Appendix A) that a time sampling frequency

$$F_s = 2c \max_{\xi:|F(\xi)|\neq 0} \{\xi\} \tag{C.28}$$

is required to recover the original signal correctly.

When such a signal has a spatial circular band shape, thus belonging to the class seen in section C.3, the relation

$$\max_{\xi:|F(\xi)|\neq 0} \{\xi\} = B \tag{C.29}$$

holds, so that Equation (C.28) simplifies into

$$F_s = 2cB\,. \tag{C.30}$$

Relation (C.30) has an immediate physical interpretation: waves having wavelength equal to $1/B$, propagating at speed c along a medium, exhibit a temporal frequency equal to cB. In order to preserve information in their sampled versions, they must be sampled above twice their frequency.

The value given by Equation (C.30), if used as the sampling frequency of a 2D-resonator model (realized by means of WMs having critical waveguide lengths), causes inaccurate positioning of the modal frequencies. Assumed that dispersion (see section C.2.1) cannot in general be eliminated, as it is a consequence of the finite number of directions a 2D signal can propagate along a WM[3], the propagation speed can at least be set to its physical value in low frequency. This can be done by simply rescaling F_s according to the geometry.

Recalling the procedure leading to Equation (C.7), we can reformulate the spatial phase shift of a signal traveling along the ideal membrane, during a time period $T = 1/F_s$. Hence, we find the following expression for the ratio (C.8), reformulated using for each geometry its respective critical waveguide length:

$$\tilde{k}_{\mathrm{g}}(\xi_x, \xi_y) = \frac{1}{2\pi\alpha_{\mathrm{g}} D\xi} \arctan \frac{\sqrt{4 - \tilde{b}_{\mathrm{g}}^2}}{\tilde{b}_{\mathrm{g}}} \,, \tag{C.31}$$

where $\tilde{b}_{\mathrm{g}}$ has the structure of b_{g} (see Equation (C.5)), but the waveguide length D has been replaced by its critical counterpart D_{g}.

Recalculation of the limit (C.10) gives:

$$\tilde{k}_{\mathrm{g}}(0) = \frac{D_{\mathrm{g}}}{D} k_{\mathrm{g}}(0) = \frac{1}{\sqrt{2}} \frac{D_{\mathrm{g}}}{D} \,, \tag{C.32}$$

that is, considering D_{s} as the reference waveguide length,

$$\tilde{k}_{\mathrm{s}}(0) = \frac{1}{\sqrt{2}} \,,\ \tilde{k}_{\mathrm{t}}(0) = \frac{\sqrt{2}}{\sqrt{3}} \,,\ \tilde{k}_{\mathrm{h}}(0) = \frac{\sqrt{2}}{3} \,. \tag{C.33}$$

This result shows that signals, under conditions of critical sampling, propagate in the WMs at different speeds, according to the geometry. The dispersion ratios can be set to unity at dc by using the temporal sampling frequencies

$$\overline{F}_{s,\mathrm{g}} = \frac{1}{\tilde{k}_{\mathrm{g}}(0)} F_s \,. \tag{C.34}$$

For the purpose of Section C.5, the dispersion ratio of a critically sampled 2D medium, adjusted to be one at dc, is called $\overline{k}_{\mathrm{g}}(\xi_x, \xi_y)$.

C.5 Performance

In order to compare the three geometries under critical sampling conditions, we show in Figure C.6 the contour plots of $\overline{k}_{\mathrm{g}}$, for a nominal spatial bandwidth $B = 1/2$ (corresponding to the circle in the figure).

It can be noted that the behaviors of the SWM and the TWM are not dramatically different in terms of average dispersion error: dispersion stands quite below

[3] Savioja and Välimäki [137] have pointed out that a signal, coming out from a TWM, can be frequency warped to reduce the dispersion error. This is made possible by the fact that $k_{\mathrm{t}}(\xi_x, \xi_y)$ can be approximated with a single–variable function $k_{\mathrm{t}}(\xi)$.

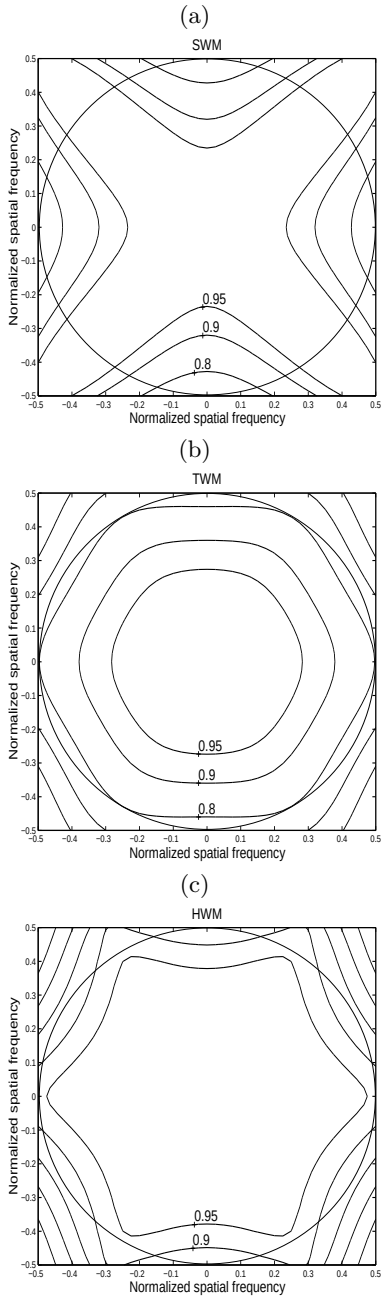

Fig. C.6. Contour plots of propagation speed ratios in the SWM (a), TWM (b) and HWM (c), when temporal sampling frequency has been set to $\overline{F}_{s,\mathrm{g}}$. All the WMs process a signal having bandwidth (depicted with the circle) equal to 1/2.

20% for most spatial frequencies in both the geometries. However, the TWM exhibits a more uniform behavior, and this uniformity can be exploited using fre-

	Waveguide Mesh			Finite Difference Scheme		
	Sq.	Tr.	Hex.	Sq.	Tr.	Hex.
additions per junction	7	11	5	4	6	3
multiplications per junction	1	1	1	1	1	1
memory locations per junction	4	6	3	2	2	2
density of junctions	1	0.866	1.732	1	0.866	1.732
density of memory locations	4	5.2	5.2	2	1.732	3.464
sample rate	1	0.866	1.5	1	0.866	1.5
additions per unit time and space	7	8.25	12.990	4	4.5	7.794
multiplications per unit time and space	1	0.75	2.598	1	0.75	2.598

Table C.1. Performance of the geometries in terms of memory requirement and computational cost (multiplications are pure bit–shifts in the SWM). Two implementations are considered: as a waveguide mesh and as a finite difference scheme.

quency warping [137]. Conversely, dispersion in the HWM stays below 10% almost everywhere, thus indicating that this geometry has the most uniform propagation speed under these test conditions.

The computational cost and the memory requirement in the different geometries can be calculated under the same conditions. They are based on Equation (C.2) which shows that each N–port lossless scattering junction requires $2N$ operations to compute the wave signals coming out from the junctions, to be stored into N locations belonging to the adjacent digital waveguides. These operations amount to $2N-1$ additions, and 1 multiplication (which can be replaced by a bit shift in fixed–point implementations of the SWM).

Table C.1 summarizes the performance, when the reference sampling rate of the square mesh $\overline{F}_{s,\mathrm{s}}$ has been set to a nominal value equal to unity. Numbers are given for two implementations: as a waveguide mesh (with memory in the waveguide branches), and as a finite difference scheme (with memory in the junctions).

The numbers of operations and memory locations per junction for the WM follow directly from Equation (C.2). The numbers of operations and memory locations per junction for the finite difference scheme follow directly from Equation (C.3). The densities of junctions result from Equations (C.21) and (C.26). The densities of locations are obtained by multiplying the third row times the fourth row. The sample rates are a consequence of Equation (C.34).

The results of table C.1 show that the triangular finite difference scheme uses the least quantity of memory. Among WM implementations, the SWM is the most efficient in terms of memory occupation.

Finally, the number of additions (multiplications) per unit time and space results by multiplication of the number of additions (multiplications) per junction, density of junctions, and sample rate. Once again, the SWM has the least density of operations when a WM model is required. Conversely, the square and the triangular geometries have about the same computational requirements in a finite difference implementation.

C.5.1 Numerical example

Let us design a WM, capable of modeling in real–time an ideal 2D round resonator, of radius $r = 0.1$ m, where waves propagate at a speed c equal to 130 m/s. Let the signal contain information up to a frequency $f = 10$ kHz.

The spatial bandwidth B of the signal traveling along the resonator is:

$$B = \frac{f}{c} = \frac{10000 \text{ Hz}}{130 \text{ m/s}} = 76.923 \text{ m}^{-1} , \tag{C.35}$$

and consequently the critical waveguide lengths required to model the signal, in the respective geometries, are:

$$\begin{aligned} D_{\mathrm{s}} &= \frac{1}{2B} = 6.5 \text{ mm} ; \\ D_{\mathrm{t}} &= \frac{1}{\sqrt{3}B} = 7.5 \text{ mm} ; \\ D_{\mathrm{h}} &= \frac{1}{3B} = 4.3 \text{ mm} . \end{aligned} \tag{C.36}$$

This means that the numbers of junctions — N_{s}, N_{t} and N_{h}, respectively — needed in the three models are[4]

$$\begin{aligned} N_{\mathrm{s}} &\approx \frac{\pi r^2}{D_{\mathrm{s}}^2} = 744 ; \\ N_{\mathrm{t}} &\approx \frac{\pi r^2}{\frac{\sqrt{3}}{2} D_{\mathrm{t}}^2} = 645 ; \\ N_{\mathrm{h}} &\approx \frac{2}{3} \frac{\pi r^2}{\frac{\sqrt{3}}{2} D_{\mathrm{h}}^2} = 1308 . \end{aligned} \tag{C.37}$$

The time sampling frequencies needed to have $\overline{k}_{\mathrm{g}}(0) = 1$ are, from Equation (C.34):

$$\begin{aligned} \overline{F}_{s,\mathrm{s}} &= 2\sqrt{2} f = 28.285 \text{ kHz} ; \\ \overline{F}_{s,\mathrm{t}} &= 2\sqrt{\frac{3}{2}} f = 24.495 \text{ kHz} ; \\ \overline{F}_{s,\mathrm{h}} &= 2\frac{3}{\sqrt{2}} f = 42.427 \text{ kHz} . \end{aligned} \tag{C.38}$$

C.6 Conclusion

A novel wave propagation model has recently been introduced for the simulation of isotropic multidimensional media. It makes use of structures called waveguide

[4] The values result by calculating the number of small areas, each one being associated with its own lossless scattering junction, which tessellate the resonator.

meshes. Even if most of the waveguide mesh properties have already been understood, there was lack of literature about their performance from a signal–sampling viewpoint.

In this paper, some properties of the most common 2D waveguide meshes — square, triangular and hexagonal — related with the bandwidth of the signal traveling on them, have been inspected. In particular, it has been shown that the triangular waveguide mesh is capable of processing a larger bandwidth than the square or hexagonal waveguide meshes having the same digital waveguide lengths.

Furthermore, when processing signals of the same bandwidth, the triangular waveguide mesh does not exhibit a computational and memory load much larger than the most computationally–efficient waveguide mesh — the square mesh — such that it can be considered, for the uniformity of its dispersion error, a good choice both in terms of simulation errors and computational cost.

The analysis presented in this article may be extended to 3D media, comparing the three geometries which most directly correspond to the waveguide meshes here reviewed: 3D rectilinear, dodecahedral, and tethraedral.

C.7 Acknowledgment

We would like to thank Lauri Savioja and Vesa Välimäki for many insightful discussions. We are also grateful to the anonymous reviewers for their constructive criticism.

C.8 Appendix - Sampling of a signal traveling along an ideal membrane

An ideal membrane establishes a relation between the spatial Fourier transforms of the signal s traveling on it, taken in correspondence of two times, $\tilde{t}$ and t:

$$S(\xi_x, \xi_y, t) = e^{j2\pi c(t-\tilde{t})\xi}\, S(\xi_x, \xi_y, \tilde{t})\ . \tag{C.39}$$

This relation has the following interpretation: each spatial component of the signal, traveling along a distance equal to $c(t-\tilde{t})$ during time $(\tilde{t}, t)$, has no magnitude variation, and has phase variation equal to $-c(t-\tilde{t})\xi$ [157].

Let us excite, at time $\tilde{t}$, the ideal membrane with a spatially band limited signal $f(x, y)$ having Fourier transform $F(\xi_x, \xi_y)$. The application of (C.39) gives the spatial Fourier transform of the signal on the membrane:

$$S(\xi_x, \xi_y, t) = \begin{cases} 0 & , t < \tilde{t} \\ e^{j2\pi c(t-\tilde{t})\xi}\, F(\xi_x, \xi_y), & t \geq \tilde{t} \end{cases} \tag{C.40}$$

From this relation, we can calculate the critical temporal sampling frequency required to sample, without loss of information, a signal traveling on an ideal membrane. In fact, the Fourier transform of the signal calculated by its spatial Fourier transform,

$$S(\xi_x, \xi_y, f) = \int_{-\infty}^{+\infty} S(\xi_x, \xi_y, t) e^{-j2\pi ft}\, dt\ , \tag{C.41}$$

has a magnitude equal to

$$\begin{aligned} &|S(\xi_x, \xi_y, f)| \\ &= \left| \int_{-\infty}^{+\infty} S(\xi_x, \xi_y, t) e^{-j2\pi ft}\, dt \right| \\ &= \left| \int_{\tilde{t}}^{+\infty} e^{j2\pi c(t-\tilde{t})\xi}\, F(\xi_x, \xi_y) e^{-j2\pi ft}\, dt \right| \\ &= \left| \int_{\tilde{t}}^{+\infty} e^{-j2\pi c\tilde{t}\xi}\, F(\xi_x, \xi_y) e^{-j2\pi t\{f - c\xi\}}\, dt \right| \\ &= |F(\xi_x, \xi_y)| \left| \int_{\tilde{t}}^{+\infty} e^{-j2\pi t\{f-c\xi\}}\, dt \right|\ . \end{aligned} \tag{C.42}$$

In particular, when $\tilde{t} \to -\infty$:

$$\begin{aligned} &|S(\xi_x, \xi_y, f)| \\ &= |F(\xi_x, \xi_y)| \lim_{\tilde{t} \to -\infty} \left| \int_{\tilde{t}}^{+\infty} e^{-j2\pi t\{f-c\xi\}}\, dt \right| \\ &= |F(\xi_x, \xi_y)| \left| \int_{-\infty}^{+\infty} e^{-j2\pi t\{f-c\xi\}}\, dt \right| \\ &= |F(\xi_x, \xi_y)|\ \delta(f - c\xi)\ . \end{aligned} \tag{C.43}$$

This means that some spectral power exists for any temporal frequency (equal to $c\xi$) associated to a spatial component of frequencies (ξ_x, ξ_y) excited in the membrane. Invoking the sampling theorem [110], the critical time sampling frequency F_s results to be equal to (C.28).

D

A Modified Rectangular Waveguide Mesh Structure with Interpolated Input and Output Points

Federico Fontana, Lauri Savioja and Vesa Välimäki
Proc. International Computer Music Conference, pages 87–90, La Habana, Cuba, September 2001.

The rectangular waveguide mesh presents aspects of redundancy when computing the impulse response of an ideal resonator. Its structure is thus modified, to define a new structure where redundant computations and unnecessary memory consumption are removed. The modified mesh saves half of the memory and computations, meanwhile it preserves all the numerical properties of the rectangular waveguide mesh accounting for stability, direction-dependent propagation speed of the wavefronts and so on. A general method, which is derived by bilinear interpolation and deinterpolation, is adapted for generalizing the input/output point positions in the modified structure. Simulations confirm the conjectures advanced herein.

D.1 Introduction

Waveguide meshes [158], in their several formulations, are deserving a certain attention by researchers concerned with the design of resonators for musical, audio and multimedia applications. Previous studies [50,138], for instance, have inspired some interesting applications of the waveguide mesh (from now denoted as WM) for modeling percussion instruments [2, 51], string instrument bodies [74], and in the analysis of reverberant enclosures [106, 134].

Although some formulations of the WM allow to realize models that closely approach the behavior of an ideal resonator, the original structure [158] remains a valuable trade-off between computational efficiency and versatility. Moreover, it allows a straightforward implementation, and this is very useful at least when one conducts preliminary tests of reliability of a WM model with respect to a given modeling problem.

In that case, minimizing the computational cost of the procedure is often an important issue. Any redundancy introduced in the computations by a heavier implementation of the model translates into unnecessary hardware usage, and

longer wait for the output. This is true especially when the WM is used in problems involving accurate modeling of huge resonators: in this case, processing time and memory consumption often become a critical factor.

The rectangular (square) WM (SWM from now) exhibits this redundancy. This can be noted, for example, by analyzing the spectrum of its magnitude response: such a response in fact mirrors at half of the Nyquist frequency, independently from where the impulse is injected and the response is taken, suggesting that half of the computations performed during the calculation of this response are unnecessary. We show that, in these cases, an SWM can be turned into a similar, lighter structure saving half of the memory and computations. In spite of this, the performance of the SWM is preserved, since only redundant information is removed by the modified structure.

Moreover, a method to calculate the output in correspondence of a point that is misaligned from the junction positions is proposed, together with a strategy for injecting a signal when the input position differs from any scattering point position. Such a generalized excitation and acquisition points can be obtained by applying bilinear interpolation to the existing scattering points in the modified SWM [136]. Though, the interpolation and deinterpolation formulas must be rewritten for accommodating them to the modified structure.

An alternative treatment of these arguments, based on the theory of Digital Waveguide Networks, will be cited whenever it has contributed to inspire this research, and when the conclusions are similar [13]. Rather, the background of the present work is more focused on the theory of audio signals.

D.2 A Modified SWM

The impulse response of an SWM is known to present redundant information, in a way that its spectrum mirrors at half of the Nyquist frequency. This characteristic does not depend on the position of the junction where the impulse is fed, and the position from where the response is taken. From a signal-theoretic viewpoint, this descends from the fact that the discrete-time transfer function which describes the response of the SWM with respect to the input and output positions is always a function of the variable z^{-2} [158]. Since this property comes from the numerical scheme which is computed by the mesh, it appears also in a realization made adopting a Finite Difference Scheme.

This issue has been outlined also by Bilbao [13], who shows that the SWM can be subdivided in two mutually exclusive schemes. Their independence comes from the fact that the SWM processes two subsets of data at each time sample, and never performs any kind of merging between them. As a result, a sample which is taken from one scattering point at a given time step does not have structural relations with the sample which is taken at the previous (or following) time step from the same point.

The excitation in general establishes a mutual dependence between adjacent samples, so that the output from the SWM is in most cases informative over the whole spectrum. In spite of this, there are situations where one of the two subsets becomes unnecessary. In particular, the impulse response comes from the

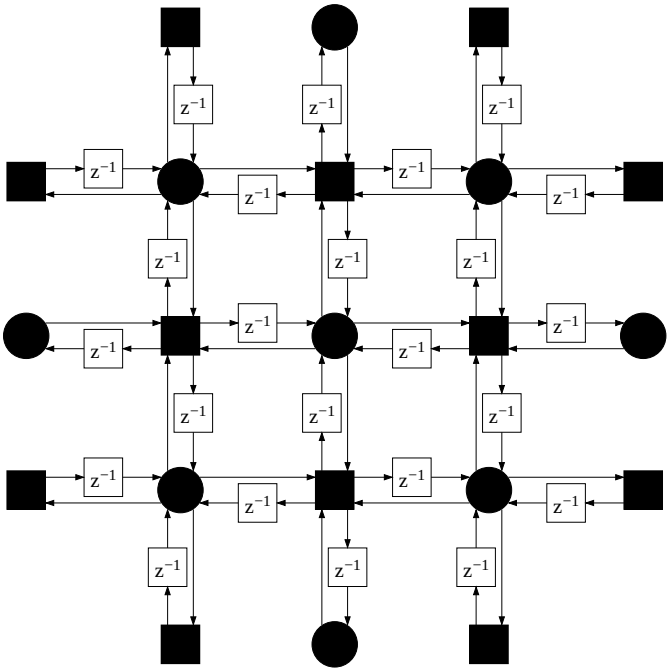

Fig. D.1. Modified SWM (size 5×5) having reflecting edges. • are scattering junctions belonging to $\mathcal{S}_{2n}$. ■ are scattering junctions belonging to $\mathcal{S}_{2n+1}$. Junctions at the boundary reverse the sign of the incoming waves.

excitation of only one of them, hence the spectral content coming from the samples belonging to the latter subset, which are always equal to zero, is redundant.

Consider an SWM as a connection of objects, each one made of one N-port scattering junction and N outgoing branches containing one unit delay[1]. Now, modify the mesh structure by interleaving these objects with new ones, where the unit delays have been substituted by delay-free connections. Denote the set containing the former objects with $\mathcal{S}_{2n+1}$, and let $\mathcal{S}_{2n}$ be the set containing the new ones. Figure D.1 depicts the structure resulting from an SWM of size 5×5 having perfect reflection at the boundaries.

At each time step, junctions belonging to $\mathcal{S}_{2n}$ compute the signal in the same way as it happens in the normal SWM. Yet, wave signals are sent without delay back to the junctions belonging to $\mathcal{S}_{2n+1}$. Hence, samples which do not keep information are overwritten with values that, in the normal SWM, are computed during the following time step (see Figure D.2).

The stability and propagation properties of the modified structure are the same as in the SWM, since the numerical scheme realized by the modified mesh includes the scheme computed by the SWM. From a signal-theoretic viewpoint, the information which is present in the modified structure equals the information conveyed by the SWM.

[1] Clearly, it will be $N = 4$ for the SWM, and $N = 6$ for the 3-D rectangular WM [134].

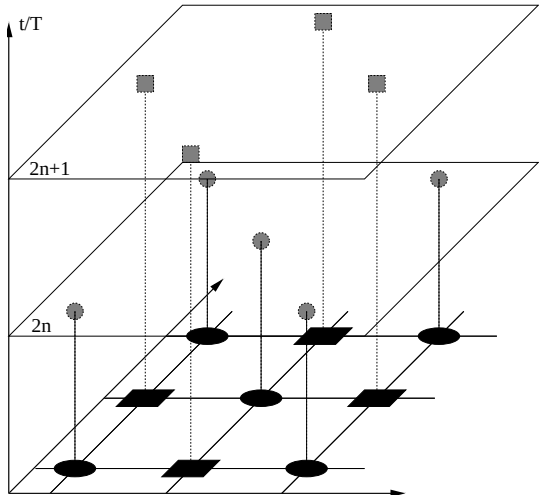

Fig. D.2. Interpretation of the signal generated by the modified SWM. Samples produced by the junctions in $\mathcal{S}_{2n+1}$ follow the samples produced by the junctions in $\mathcal{S}_{2n}$.

The modified SWM is convenient whenever the excitation signal mirrors at half of the Nyquist frequency. Under this assumption, the global amount of memory in the new structure is halved, while the number of operations needed to calculate the output is in principle the same. Though, considering that two time samples of the output are calculated during each computation cycle, the number of operations that are needed to compute the output signal is reduced by a factor of two as well. Such a convenience appears in an FDS realization of the modified scheme as well. In this case, the reduction in memory consumption appears in the form of just one unit delay associated with each scattering point, instead of two.

Simulations conducted over the two models confirm that the spectral redundancy present in the magnitude response of the SMW (Figure D.3) is removed by the modified structure (Figure D.4).

D.3 Interpolated Input and Output Points

The modified structure can be excited in correspondence of the junctions belonging to $\mathcal{S}_{2n+1}$. Likewise, each one of these junctions presents the impulse response calculated in correspondence of its own position. In spite of this, there are cases where the excitation or acquisition points cannot correspond with the positions of the scattering points.

In these cases, first-order Lagrange interpolation can be used to interpolate between junctions. This method can be extended to the two-dimensional case in the form of bilinear interpolation, whose versatility and ease of use together with the SWM have been previously shown [136].

In that treatment, the formulas accounting for coefficients w_{ij} that weight the samples s_{ij} of the (four) nearest scattering points, namely 11, 12, 21, and 22, are

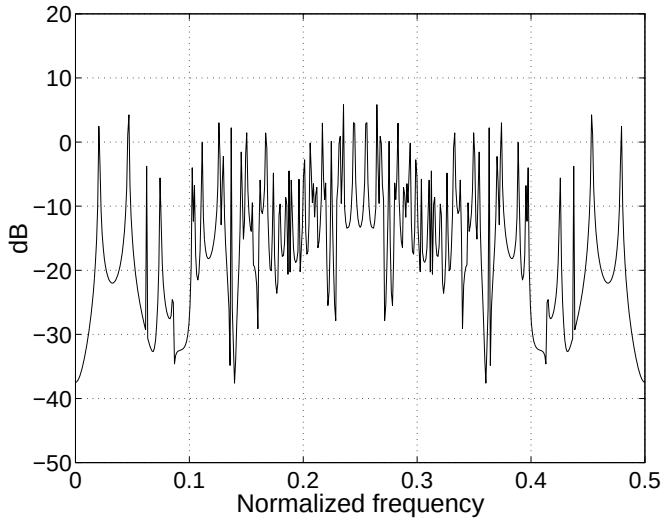

Fig. D.3. Frequency response of a SWM (size 25×25). Excitation at the center.

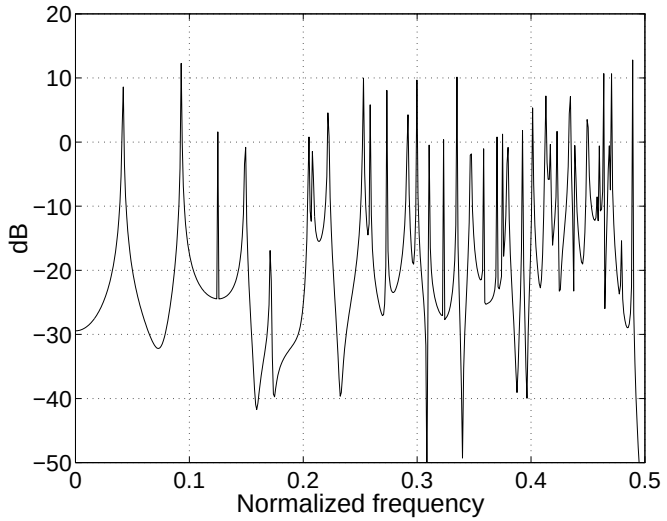

Fig. D.4. Frequency response of a modified SWM (size 25×25). Excitation at the center.

the following ones:

$$w_{11} = (1-x)(1-y)\ ,\ w_{12} = x(1-y)$$
$$w_{21} = (1-x)y\ ,\ w_{22} = xy$$

where (x, y) are the coordinates of the interpolated output v relative to position 11:

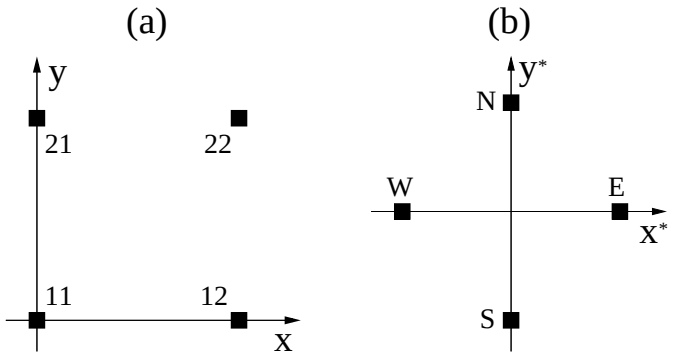

Fig. D.5. Reformulation of bilinear interpolation from original (a) to new (b) coordinates obtained by translating, rotating and stretching the old set.

$$v(x,y) = \sum_{i=1}^{2}\sum_{j=1}^{2} w_{ij} s_{ij}$$

The same weighting parameters are used to deinterpolate the input $u(x,y)$ over the nearest scattering points:

$$s_{ij} = w_{ij}\, u(x,y)$$

All these formulas hold when the digital waveguide lengths have been normalized to unity.

Such relations apply also to the junctions belonging to $\mathcal{S}_{2n+1}$ in the modified structure, once they are reformulated (see Figure D.5) in new coordinates (x^*, y^*). This transformation consists of a translation by $(-1/2, -1/2)$ followed by a rotation by $\pi/4$. The new coordinates are finally stretched by a factor equal to $\sqrt{2}$. New weighting parameters are thus found out, with the notations depicted in Figure D.5(b):

$$\begin{aligned} w_N &= (1 - x^* + y^*)(1 + x^* + y^*)/4 \\ w_W &= (1 - x^* + y^*)(1 - x^* - y^*)/4 \\ w_E &= (1 + x^* - y^*)(1 + x^* + y^*)/4 \\ w_S &= (1 + x^* - y^*)(1 - x^* - y^*)/4 \end{aligned}$$

Suppose, for example, to swap $\mathcal{S}_{2n+1}$ and $\mathcal{S}_{2n}$ in the square structure that produces the spectrum depicted in Figure D.4; we obtain a new, square mesh that is centered around a junction belonging to $\mathcal{S}_{2n}$, like the one depicted in Figure D.1. Clearly, in this case an excitation at the center can no longer be applied. If we excite the center of the mesh using deinterpolation over the four neighbors, a response is obtained like the one (in solid line) in Figure D.6, where the non-deinterpolated excitation (refer to Figure D.4) is repeated (in dashed line) for ease of comparison. Note that some amplitude distortion appears, coming from Lagrange interpolation. In this case, distortion is at its maximum since we are

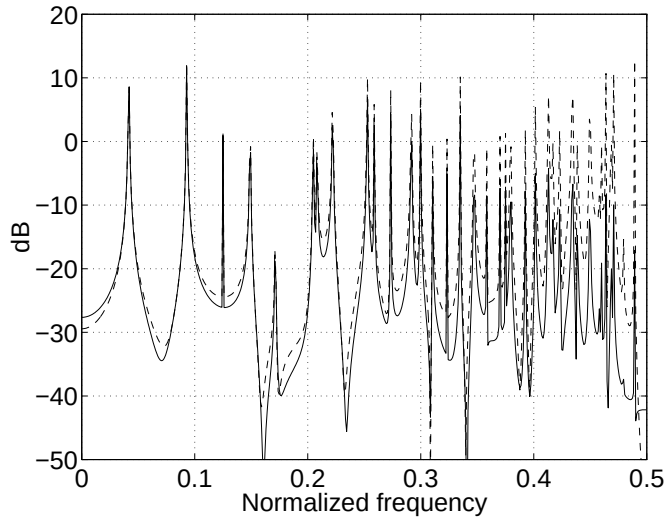

Fig. D.6. Spectrum from deinterpolated excitation around the center of a modified SWM sized 25×25 (solid line) compared with spectrum depicted in Figure D.4 (dashed line).

interpolating on a point which is farthest from any neighbor scattering junction. In general cases amplitude distortion is less severe.

Finally, note that (de/)interpolation can be applied, more in general, to any difference scheme whose scattering nodes do not match with the excitation/acquisition point positions.

D.4 Conclusion

A modified rectangular waveguide mesh has been presented. It processes a numerical scheme which is included by the square waveguide mesh. It saves half of the computations and memory without loosing any information during the calculation of the impulse response. In the case that the input and output points differ from the junction positions, bilinear deinterpolation and interpolation can be respectively applied to feed the mesh and acquire its response.

The analysis presented in this paper can be easily extended to 3-D rectangular waveguide meshes and to finite difference schemes.

E

Acoustic Cues from Shapes between Spheres and Cubes

Federico Fontana, Davide Rocchesso and Enzo Apollonio
Proc. International Computer Music Conference, pages 278–281, La Habana, Cuba, September 2001.

Solids of different shapes resonate according to their peculiar geometry. Although physics, for some fundamental shapes such as the cube and the sphere, provides explicit formulas for determining the modal frequencies, a general resonance analysis of 3-D shapes must be conducted using numerical methods. In this paper, Waveguide Meshes are used to model intermediate geometries between the cube and the sphere. This example is paradigmatic of the general problem of morphing 3-D shapes.

We are interested in understanding how smooth shape transitions from sphere to cube translate into a migration of the resonating modes. This work is aimed at assessing the suitability of the waveguide mesh as a tool for such a research, and it is preliminary to further investigations involving human subjects.

E.1 Introduction

Acoustic rendering is an emerging research field, whose growth is stimulated by two, somehow opposite factors: an increasing interest for applications of multimodal virtual reality on one side, and inevitable constraints in cost and technology of the equipment on the other side. These factors, together, lead to looking for rendering methods which are realistic and computationally efficient at the same time. Under this assumptions, a method which is capable to acoustically render objects, or enclosures, will move the listener to virtually experience a scenario without reproducing all its characteristics.

Important studies have already been conducted in the visual field, and remarkable results have been achieved in rendering shadows, textures, lightings, and object movement. This encourages to looking for audio counterparts of those methods.

Perception of shapes from acoustic cues is a matter of investigation for researchers in psychophysics [92] and object modeling [127]. A listener could experience, by hearing, to stay for example in the middle of a semi-spherical enclosure,

or in front of a large cube, without seeing any of them [96]. Tests have been conducted to investigate whether or not listeners discriminate simple shapes such as cubes and spheres [127]. Such experiments show that shape labels can be reliably attached to sounds, regardless of their pitch, and that the distribution of low-frequency resonances play the prominent role in this task. As a side result [128], it was shown that non musicians tend to equalize pitches of resonators as if they resulted from equal-volume shapes. This gives us a simple criterion that can be used to minimize the influence of pitch when experimenting with 3-D objects of varying shapes.

In this work we investigate on the "spectral continuum" holding when a cube morphs into a sphere, through specific geometries called superquadrics [88]. We look for specific cues to check out if there are "footprints" which acoustically label each intermediate shape, and, hence, the whole morphing process. We will show that these cues exist, although only listening tests could demonstrate a real sensitivity of human beings to shape variations.

A numerical method is needed for studying the ellipsoidal shapes. In our work, all resonators are modeled using waveguide meshes [158]. In particular, the 3-D triangular waveguide mesh (3DTWM) has been adopted, for its low dispersion characteristics and good approximation in modeling boundaries [54].

All the simulations have been conducted working with an application written in C++, whose user interface can be seen from the screenshot in Figure E.1. This application simplifies the construction and initialization of the mesh, and performs all needed processing. A pre-release of the executable program is available from the Web site of the SOb European Project (`http://www.soundobject.org`) for public experimentation.

E.2 From Spheres to Cubes

One possible morphing from spheres to cubes can be easily realized if we restrict any possible geometry to be an ellipsoid. Superquadrics are, in this sense, a versatile family which is defined by the following equation [88]:

$$\left|\frac{x}{a}\right|^{\gamma_x} + \left|\frac{y}{b}\right|^{\gamma_y} + \left|\frac{z}{c}\right|^{\gamma_z} = 1 \tag{E.1}$$

Changes in shape are then performed by varying only the three parameters $\gamma = \gamma_x = \gamma_y = \gamma_z$ together, and constraining a, b, c to condition $a = b = c$:

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

We consider six shapes, including the sphere and the cube, which are built according to (E.1). Their positive section (i.e., their volume limited to $\{(x, y, z) : x > 0\,,\, y > 0\,,\, z > 0\}$) is depicted in Figure E.2, where, starting from left above, parameter γ has been set to

$$\gamma_1 = 2 \;\gamma_2 = 2.2 \;\gamma_3 = 2.5$$
$$\gamma_4 = 3 \;\; \gamma_5 = 4 \;\; \gamma_6 = 10$$

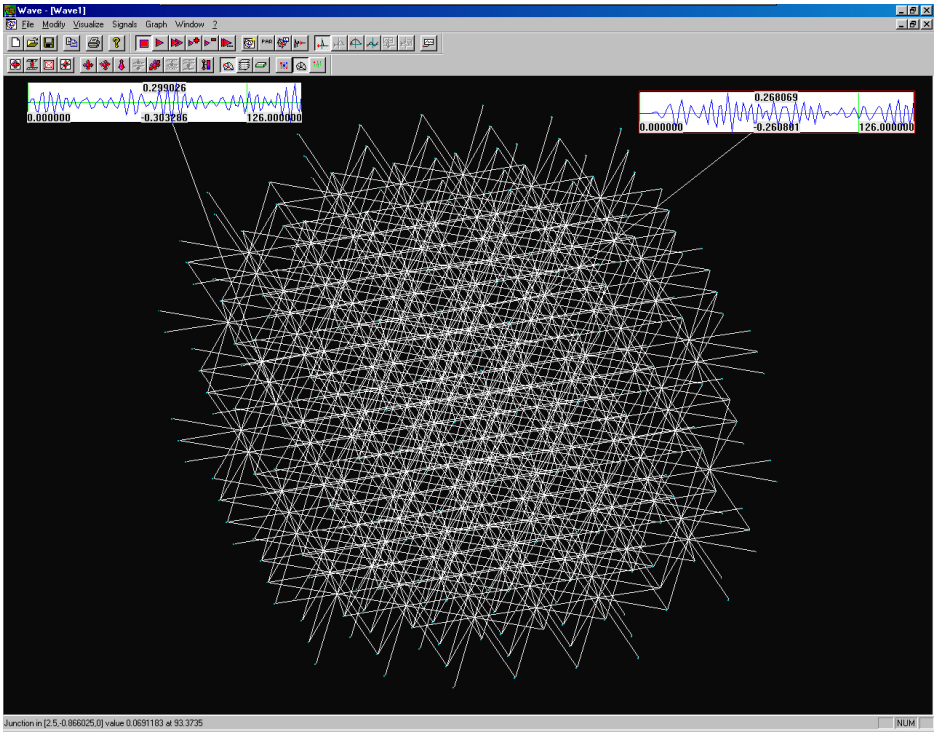

Fig. E.1. User interface of the application running the mesh models.

respectively. Note that γ_6 is large enough to represent the cube. This is true because the discrete boundary, which will be modeled by the 3DTWM, does not change for $\gamma \geq 10$.

Accordingly, waveguide mesh models are built. Taking advantage from the application presented above, 3DTWM's are constructed so that they match, as close as possible, the geometries coming from the selected ellipsoids. Perfect reflection of the signal holds at the boundary.

Figure E.3 depicts orthogonal projections to the z-plane of the mesh models, in the same order given in Figure E.2, for intermediate shapes.

Inevitable mismatching between ideal geometries and 3DTWM models can be noted, even if the scattering junction density was guaranteed to provide a minimum distance equal to 20 junctions between surfaces located on opposite sides.

Moreover, the 3DTWM models flat surfaces with lower accuracy due to its own topology. This causes some blur in the definition of the resonance peaks, especially in the case of the cube. This will be evident in the spectral analysis.

E.3 Spectral Analysis

For each geometry, one spectrum was calculated from a signal obtained by exciting (using an ideal impulse) the resonator on three points, i.e., the center plus two

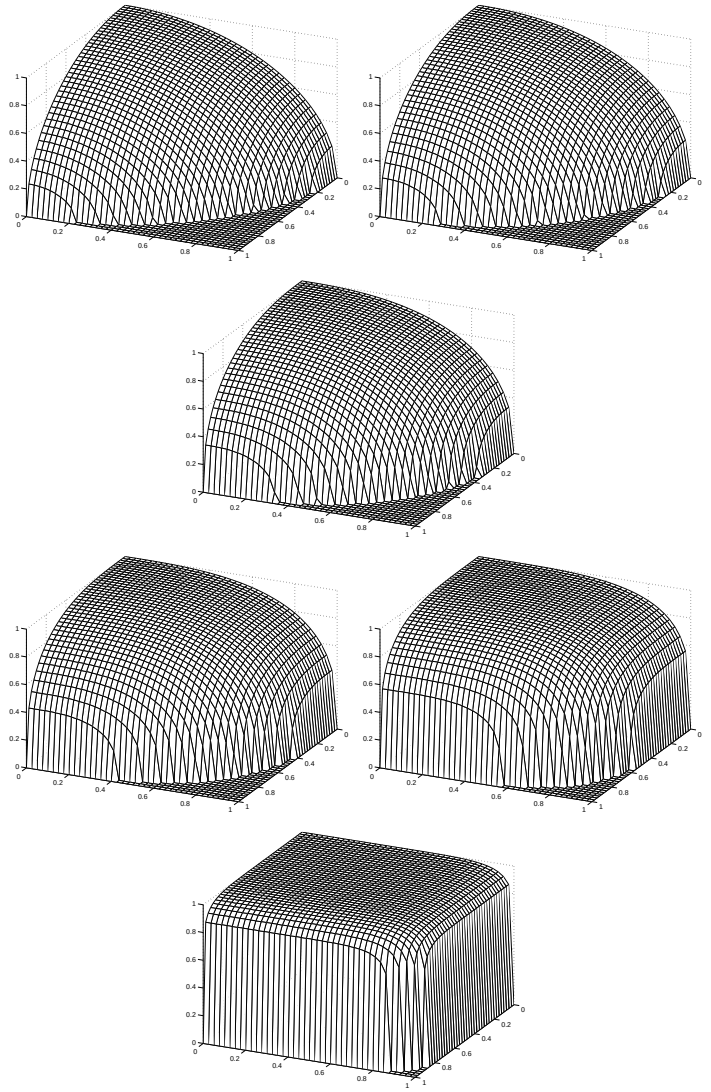

Fig. E.2. Positive sections of superquadrics obtained using (E.1). γ has been set to $\{2\,,\,2.2\,,\,2.5\,,\,3\,,\,4\,,\,10\}$ starting from left above, respectively.

points close to the boundary. The output signal was picked up on a position which was located near the boundary, for capturing most of the resonances. For non-spherical shapes, the excitation and output junctions were located near one corner. Since the corner geometry varies together with shape, inevitable variations in the excitation and acquisition positions occur during the experiment. For this reason, the dynamics of the output signals varies with shape.

Signals have been damped offline. Offline damping ensures that the resonances do not move due to imperfect internal modeling of attenuation mechanisms.

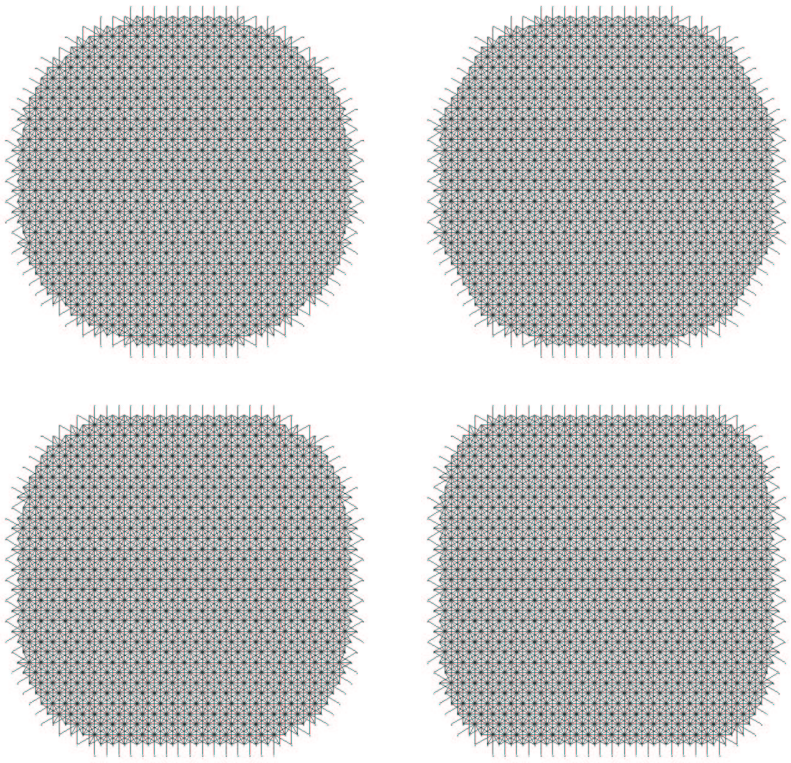

Fig. E.3. Projections, orthogonal to the z-plane, of 3DTWM models closely matching the intermediate shapes given in Figure E.2. γ has been set to $\{2.2\,,\ 2.5\,,\ 3\,,\ 4\}$ starting from left above, respectively.

Each spectrum should be rescaled in the frequency axis, holding the condition of volume constancy. Table E.1 (second column) shows, according to the chosen geometries, volume ratios for resonators having the same size[1] in the sense of Figure E.2. In the third column, size ratios for resonators having the same volume are shown. Clearly, frequency rescalings for constant volume morphing should comply with the values in column three[2]. Volume normalization of signals will be used in future research, devoted to investigate the perceptual aspects of shape variations.

We can analyze a portion equal to 1/16 of the band of the output signals. Once the sampling frequency has been set to a nominal value of 8 kHz, frequencies up to 250 Hz are hence taken into account.

Figure E.4 shows plots of the spectra discussed above, from the sphere (top) to the cube (bottom). Frequencies are expressed in Hz, and gains in dB. The

[1] say, diameter of the sphere, and side length of the cube

[2] Note that an alternative approach for obtaining resonators of equal volume would consist in modeling each geometry using meshes having, more or less, the same number of scattering junctions. This way is in practice harder to follow than numerically computing the ratios presented in table E.1.

γ_i	$\mathrm{Vol}(\gamma_i)/\mathrm{Vol}(\gamma_1)$	$\mathrm{Size}(\gamma_i)/\mathrm{Size}(\gamma_1)$
2	1	1
2.2	1.0923	0.970
2.5	1.2086	0.940
3	1.3560	0.900
4	1.5406	0.865
10	1.8157	0.815

Table E.1. Volume ratios for equal sizes (second column), and size ratios for equal volumes (third column). Geometries given by $\gamma_1, \ldots, \gamma_6$.

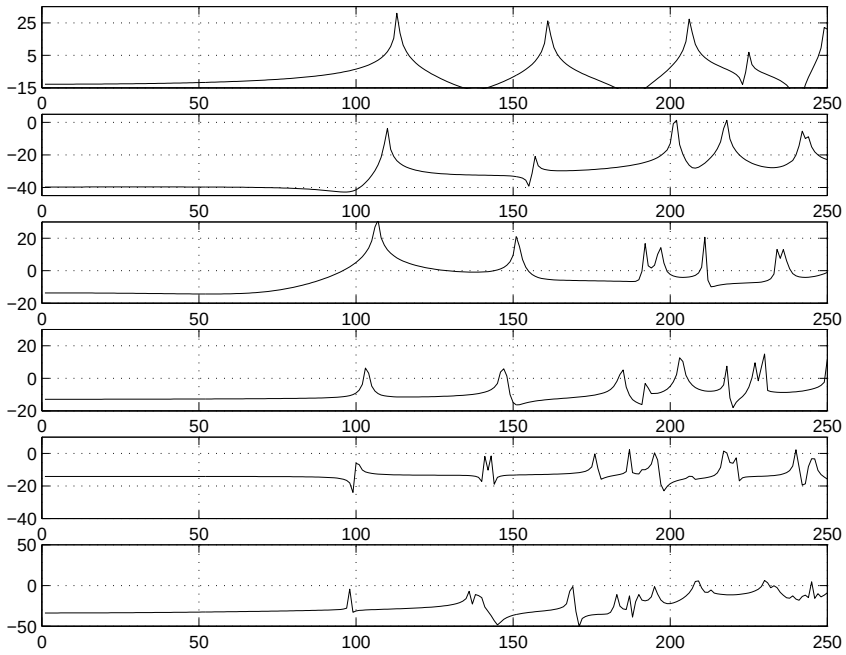

Fig. E.4. Low-frequency portion (1/16 of the nominal band) of the output signals taken from the resonators. Top: sphere. Bottom: cube. x-axis: frequency (Hz). y-axis: gain (dB).

lower resonances exhibited by the sphere are sparser, as one would expect from theory [100], and define a clear mode series where each mode accounts for a precise portion of the whole band.

As shape morphs to squareness, the mode series shrinks and shifts to lower frequencies. At the same time, new resonances arise in between the existing ones, so that the density of modes increases as the resonator approaches the cubic shape.

Both shifting of the modes to lower frequencies, and rising of new modes in between, are events which do not depend on the excitation and listening positions. Of course, these positions determine the modes which appear on the spectra or, likewise, the resonances which are audible.

From the previously shown plots, we can extrapolate a clear evolution of the low-frequency modes. This kind of description is especially useful when someone wants to recreate the low-frequency resonances of a shape such as the ones presented here, for instance using additive synthesis, or render a shape by processing a signal with a series of tunable (second-order) equalization filters, whose peak frequencies follow the positions of the modes during changes in shape.

These characteristics of the spectra have a counterpart in the sound samples which are obtained from the corresponding signals. As a resonator approaches the shape of a cube, pitch becomes less evident (or more ambiguous) and, at the same time, a sense of growing brilliance in the sound is experienced by the listener. Although pitch, if sensed, decreases as shape migrates to squareness, a proper resampling of the signals which respects volume constancy should equalize the pitch and balance the brightness to a homogeneous value.

E.4 Conclusion

A study on the spectral modifications of sounds produced by resonators, whose geometry morphs from a spherical to a cubic shape, has been presented. It has been shown that a "spectral continuum" exists such that listeners could in principle be sensitive not only to roundness and squareness, as shown by previous results, but also to certain intermediate situations.

Listening tests using sounds whose spectra are properly rescaled, will verify whether the spectral cues provided by such shapes have a perceptual counterpart.

F

Recognition of ellipsoids from acoustic cues

Federico Fontana, Laura Ottaviani, Matthias Rath and Davide Rocchesso
Proc. Conference on Digital Audio Effects (DAFX-01), pages 160–164, Limerick, Ireland, December 2001.

Ideal three-dimensional resonators are "labeled" (identified) by infinite sequences of resonance modes, whose distribution depends on the resonator shape. We are investigating the ability of human beings to recognize these shapes by auditory spectral cues. Rather than focusing on a precise simulation of the resonator, we want to understand if the recognition takes place using simplified "cartoon" models, just providing the first resonances that identify a shape. In fact, such models can be easily translated into efficient algorithms for real-time sound synthesis in contexts of human-machine interaction, where the resonator shape and other rendering parameters can be interactively manipulated. This paper describes the method we have followed to come up with an application that, executed in real-time, can be used in listening tests of shape recognition and together with human-computer interfaces.

F.1 Introduction

Recently, research topics in the field of psychophysics have been concerned with the faculty of human beings to *hear the shape*, both in the two-dimensional (2-D) and three-dimensional (3-D) case [89]. This means, for example, that sounds coming from square rather than circular membranes after an excitation, or resonances that are produced by cubic rather than spherical empty cavities, containing a sound source in their interior, may convey to the listener cues accounting for the shape of the resonator.

Several results [89, 127] seem to testify that, in the case of 3-D shapes, a fundamental role in this type of recognition is played by the spectral content of the sounds. Since, in the case of ideal resonators, the sequence of resonance modes depends only on the resonator shape, it makes sense hypothesizing that 3-D resonators convey perceptually relevant cues that are in strong correlation with their shapes.

The way these cues are perceived by the listener is a matter of investigation for ecological psychologists [61]. We decided to focus our attention in the spectral *mode series* such as a *label* of the resonator, meanwhile taking care of preserving as far as possible the constancy of all other physical and geometrical parameters. In particular, a variation of the resonator size leads to a proportional shift in frequency of the mode series: a solution must be found to govern those shifts during changes in shape of the resonator or, in other words, a rule to infer the size must be found once a shape is given.

The simplest idea would be constraining the fundamental mode to a unique value during changes in shape. This approach is quite non-ecological. Rather, a psychophysically more well-founded rule suggests to *preserve the constancy of the resonator volume* during changes in shape [128]: following this approach, the fundamental mode shift is minimal for intermediate shapes between the cube and the sphere.

A rule is needed to map one or more *morphing* parameters into corresponding shapes. Superquadrics [88] have been adopted here to realize direct and versatile maps: using just one parameter, geometries that are consistent with the problem can be selected via a simple set of equations. These geometries can be used to initialize models of resonators, if these models can be directly "shaped" exactly like the resonators should be. Waveguide meshes [54,158] comply with this requirement, and represent a good modeling solution, with several pros and cons that have been explained in more detail in previous literature [55].

In particular, Waveguide meshes allow to select the excitation and acquisition points in the resonators. In this way, the mode series can be detected in regions where the modal density is particularly rich, and, conversely, in regions that are *nodal* with respect to many resonances [48]. Since our investigation needs only to deal with the first part of the mode series (i.e., few tens of modes), the waveguide models allow to assess with enough precision all the resonances that are present in the mode series: this is done detecting the signal on acquisition points that are located near the corners of the resonator. Alternatively, for reasons of symmetry and on a practical and ecologically-consistent basis, the center has been chosen such as the region where only some resonances are audible.

Smooth changes in shape using the morphing parameter translate into progressive changes in the resonances positions. In the meantime, smooth changes in the acquisition point, moving between the center and the corner of the resonator, translate into corresponding variations in amplitude of the resonance peaks. Hence, the controls of shape and position map into intuitive features of the frequency responses.

All these features can be easily reproduced using a filter bank of second-order filters, where each filter is tuned to one particular resonance frequency [98]. Moreover, this filter bank has a precise and physically meaningful interpretation [7]. Simple methods like linear interpolation can be used to interpolate between responses that have not been simulated.

We have developed a pd-module [115] that implements such a filter bank. It is controlled in the parameters of shape (between sphere and cube), and listening points (between center and corner). Using this module, proper sounds can be convolved as if they were listened from a point located in a 3-D cavity having

a given shape. This module realizes a so-called *cartoon* model [62], that can be used in interactive real-time environments. In particular, we are going to use it in listening tests of shape recognition.

F.2 Geometries

One possible morphing from spheres to cubes can be realized if we constrain the geometries to be ellipsoids. Superquadrics are, in this sense, a versatile family of ellipsoids. For our purpose, we restrict their use to a set of geometries that is defined by the following equation in the 3-D space [88]:

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

Changes in shape are simply determined by varying γ, that acts like a morphing parameter:

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

We have analyzed eight shapes, including the sphere and the cube, which have been built according to (F.1). The corresponding values of the morphing parameters are the following ones:

$$\gamma_1 = 2 \; \gamma_2 = 2.2 \; \gamma_3 = 2.5 \; \gamma_4 = 3$$
$$\gamma_5 = 4 \; \gamma_6 = 6 \quad \gamma_7 = 20 \; \gamma_8 = 100$$

γ_8 results in a geometry that approximates the cube with good precision. Positive sections (i.e., volumes limited to $\{(x, y, z) : x > 0\,,\, y > 0\,,\, z > 0\}$) of some of these geometries (γ_3, γ_4 and γ_5) are depicted in Figure F.1, starting from left.

Then, 3-D Waveguide mesh models have been designed, in such a way that they reproduce ideal resonators having a shape that matches, as close as possible, the geometry given by the corresponding superquadric. Figure F.2 depicts, for the same geometries and with the same ordering seen in Figure F.1, orthogonal projections of the mesh models that have been used in this context. The reader can find further details of the modeling strategies that have been adopted in this research in a previous paper [55].

F.3 Simulations

For each chosen geometry two impulse responses have been computed. The resonator model was fed with energy in a way that all the modes in the scope of our analysis were excited[1]. The responses were acquired from junctions located near the corner and at the center. In this way two mode series were collected from each shape, one accounting for all the modes that a resonator can produce in the lower

[1] in practice this required to feed several junctions of the mesh with an impulse.

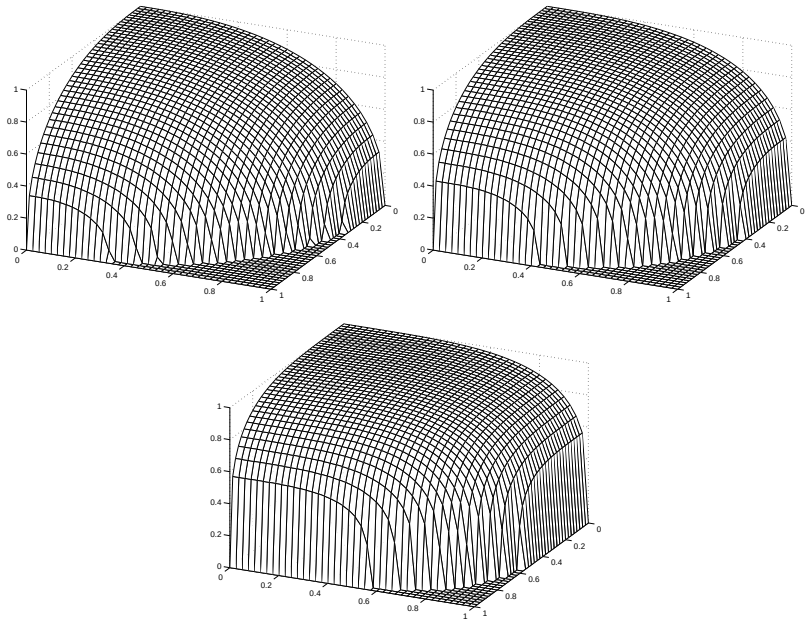

Fig. F.1. Positive sections of geometries obtained using (F.1). γ has been set to $\{2.5\,,\,3\,,\,4\}$, starting from left.

part of the frequency domain, the other presenting only the modes resonating at the center of a 3-D ideal resonator, respectively.

Since changes, both in shape and in the acquisition position, do not introduce discontinuities, the corresponding responses exhibit smooth and continuous variations as well. Changes in the acquisition point result in mode cancellations that depend on the nodal regions falling on that point. Such cancellations are present in the spectra in the form of missing resonances.

The effect of changes in shape is more complicate: they result not only in shifts of the modes, but also in resonance splits and merging. By this phenomenon, the higher mode density in the case of the cube turns to be possible.

Figure F.3 depicts all the responses that have been calculated in our analysis, both acquiring the signal at the center (dashed line) and at the boundary (solid line). Parameters $\gamma_1, \ldots, \gamma_8$ are ordered starting from above. All frequency domains are in Hz, and gains are in dB. Each resonator has been resized to maintain the volume constant with shape.

The fundamental frequency value is uninfluential in our experience. It has been set in the sphere to a nominal value of 415 Hz. Other fundamental frequencies are constrained by volume constancy: in practice, they slightly move with shape toward a lower frequency [126].

Mode canceling due to changes in the acquisition point are evident for all the geometries. It can be interesting to notice that the canceled modes do not change with shape, so that the responses taken from the center exhibit an overall

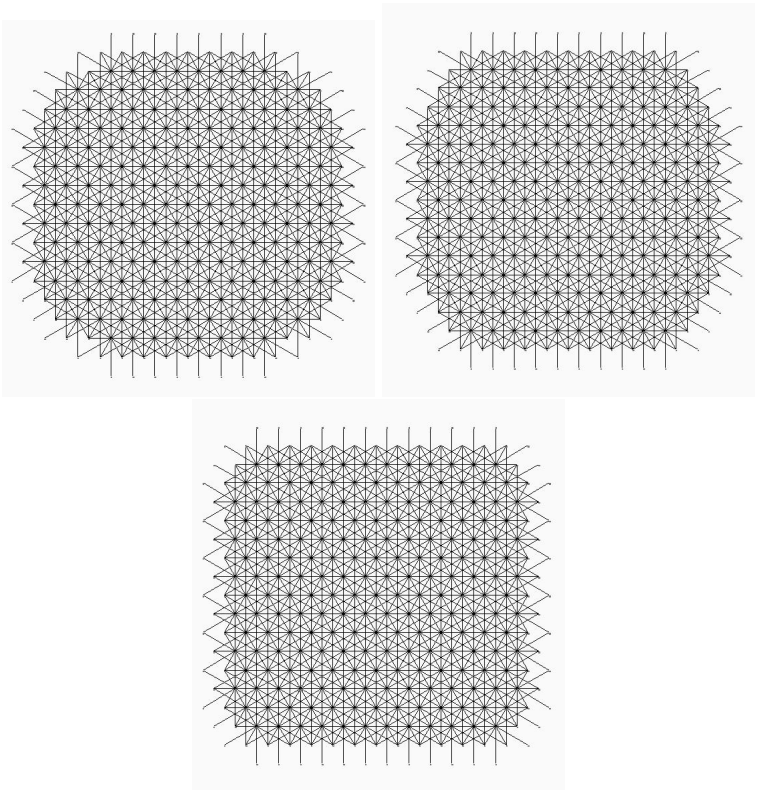

Fig. F.2. Orthogonal projections of Waveguide meshes modeling the geometries seen in Figure F.1. γ has been set to $\{2.5\,,\,3\,,\,4\}$, starting from left.

homogeneity of behavior. A definite homogeneity is quite evident also for the responses that are acquired near the boundary. Mode splitting and merging is figured out in particular focusing on the very first modes.

Finally, the theoretical mode positions (depicted in Figure F.3 with '$\circ$' for the sphere, and with '$\times$' for the cube), as they are calculated using analytical tools, match well with the resonances computed by the simulations. This suggests a correct use of the Waveguide mesh models in our research.

F.4 Cartoon models

Transfer functions having magnitude plots such as the ones shown in Figure F.3 can be realized straightforwardly. The most versatile and probably best-known solution makes use of second-order tunable equalization filters [119]. In spite of this we have adopted, like in the case of the volume constancy rule seen in Section F.3, a solution that has a more consistent physical background, although respecting the requirement of efficiency and versatility.

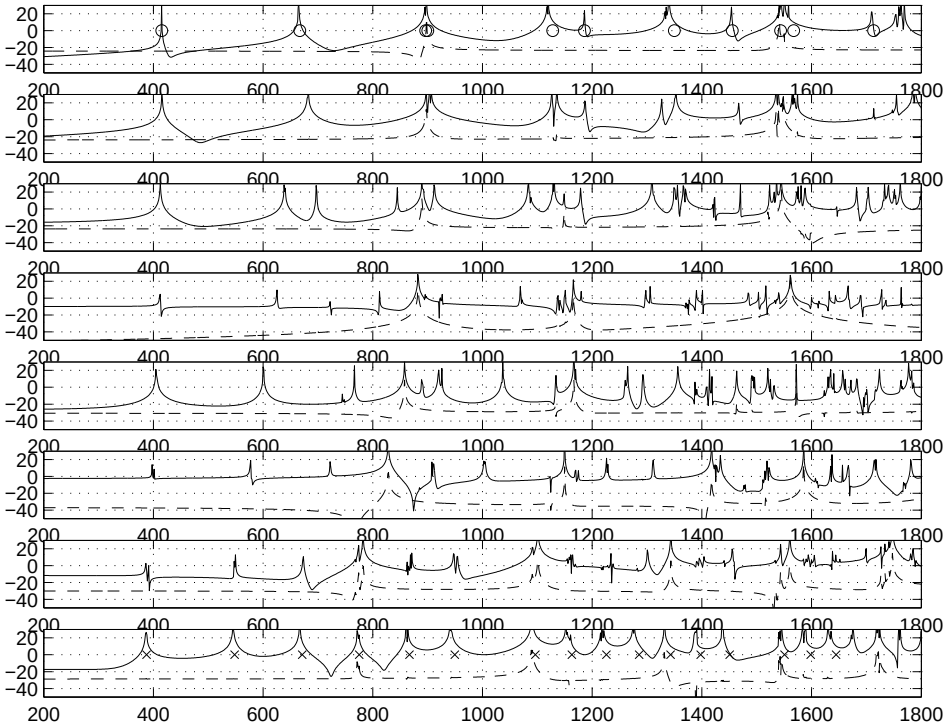

Fig. F.3. Plots of frequency responses from resonators defined by morphing parameters $\gamma_1, \ldots, \gamma_8$, starting from above. Solid line: acquisition near the boundary; dashed line: acquisition at the center. All frequency domains are in Hz, gains are in dB. Theoretical positions of the resonances in the sphere are depicted with '$\circ$'; theoretical positions of the resonances in the cube are depicted with '$\times$'. All resonators resized to maintain the volume constant with shape. Nominal frequency of the fundamental mode in the sphere has been set to 415 Hz.

Given the first N resonance modes generated by an ideal resonator, we can reproduce them using a one-dimensional (1-D) physical system consisting of N elementary blocks in series, each one being made of one mass and one spring. A damper is added to each block to provide a lossy component, giving physical consistency and realism to the model. In this way each elementary block independently governs the corresponding mode. More precisely, the parameters of frequency position and decay time of a mode are computed by simple functions of the mass, the spring constant and the damping factor [7]. After discretization, we obtain a physical model of a 1-D resonator in the form of a parallel second-order filter bank, where each filter in the bank accounts for a single mode of the resonator.

The 1-D resonator is, so, a cartoon model of the cavity [62]. Although justified from a physical modeling viewpoint, this model also allows a quite straightforward control of the position and the amplitude of each mode. As seen in Section F.3, these two parameters can be seen as resulting from a particular choice of the resonator shape and sound acquisition point. Hence, we can think to set up a rule that maps couples of (shape,position) into couples of (frequencies, amplitudes):

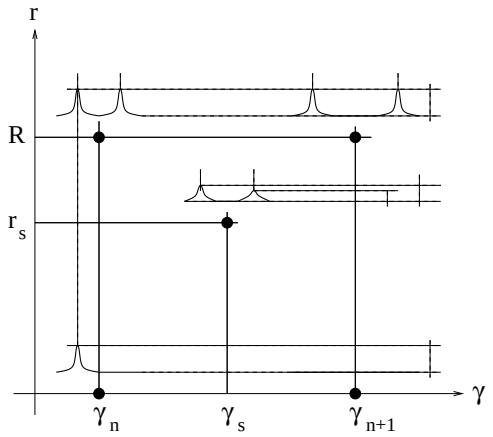

Fig. F.4. Lagrange interpolation of gains and frequency positions.

$$(\gamma, r) \longrightarrow (\boldsymbol{\omega}, \boldsymbol{G}) = ([\omega_1, \ldots, \omega_N], [G_1, G_N]) \tag{F.2}$$

where r is the distance of the acquisition point from the center, measured along a pre-determined direction common for all shapes ($0 \leq r \leq R$, R being the distance between the boundary and the center), $\omega_i, i = 1 \ldots N$ are positions in frequency and $G_i, i = 1 \ldots N$ are gains of the N modes at the acquisition point; γ, of course, selects the shape ($\gamma_1 \leq \gamma \leq \gamma_8$).

A careful design of such a map would require the knowledge of several responses from each geometry, since the nodal regions combine in a wide variety over different acquisition points. Moreover, a precise reproduction of the modes may complicate the map expressed by (F.2) up to a point where a control in real time of the 1-D model could in principle become difficult. For this reason, we have "cartoonified" also the control layer, linearly interpolating between couples (γ, r) where the image $(\boldsymbol{\omega}, \boldsymbol{G})$ is not known.

Suppose to set the input parameters to (γ_s, r_s), such that the N mode positions and amplitudes, $(\boldsymbol{\omega}_s, \boldsymbol{G}_s)$, require interpolation. *Bi-linear* (Lagrange) interpolation requires to calculate $\boldsymbol{\omega}_s$ and $\boldsymbol{G}_s$ using relations that involve four interpolated points, where the mode positions and amplitude are known, and the distance between these points and the interpolation point. If (see Figure F.4) the interpolated points are respectively labeled with $(n, 0)$, (n, R), $(n+1, 0)$, $(n+1, R)$ (n is a number between 1 and 7), such relations become:

$$\begin{aligned}
\boldsymbol{\omega}_s &= \boldsymbol{\omega}_n + \frac{\gamma_s - \gamma_n}{\gamma_{n+1} - \gamma_n}(\boldsymbol{\omega}_{n+1} - \boldsymbol{\omega}_n) \\
\boldsymbol{G}_s &= \frac{R - r_s}{R}\frac{\gamma_{n+1} - \gamma_s}{\gamma_{n+1} - \gamma_n}\boldsymbol{G}_{n,0} + \frac{r_s}{R}\frac{\gamma_{n+1} - \gamma_s}{\gamma_{n+1} - \gamma_n}\boldsymbol{G}_{n,R} \\
&\quad + \frac{R - r_s}{R}\frac{\gamma_s - \gamma_n}{\gamma_{n+1} - \gamma_n}\boldsymbol{G}_{n+1,0} + \frac{r_s}{R}\frac{\gamma_s - \gamma_n}{\gamma_{n+1} - \gamma_n}\boldsymbol{G}_{n+1,R}
\end{aligned}$$

Note that bi-linear interpolation reduces to linear interpolation between two points in the first equation. In fact, mode positions are independent from the acquisition point.

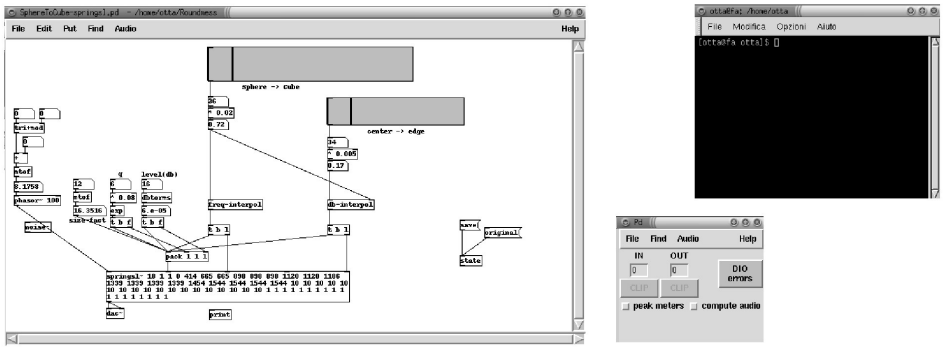

Fig. F.5. Implementation of the model as a *pd module*: a snapshot of the graphic user interface.

F.5 Model implementation

The described model has been implemented as a *pd module* under a PC running Linux and pd [115]. After acquiring a sound, it performs a convolution introducing the selected resonances in the audio signal. The simplicity of the module results in a low computational load for the hardware, so that enough resources are left to the graphic interface.

Figure F.5 shows a snapshot of the module interface, that provides an interactive, real-time environment where the user can modify the parameters of shape and the acquisition point, receiving an immediate auditory feedback from the system.

This implementation is going to be used in listening tests of shape recognition. During these tests, particular care must be taken in the choice of the sound that will be processed by the filter bank, since poor or inadequate sounds can result in inconsistent tests. Scope of these tests is investigating if the auditory cues that are conveyed by our model can evoke sensations of shape, or improve the effectiveness of a multi-modal display devoted to convey information about 3-D shapes. For this reason, the model is going to be implemented also in a system for human-machine interaction involving a more sophisticate interface.

F.6 Acknowledgments

This research has been supported by the European Commission under contract IST-2000-25287 ("SOb - the Sounding Object"). `http://www.soundobject.org`.

The Authors are grateful to Mr. Federico Avanzini for his helpful suggestions.

G

A Structural Approach to Distance Rendering in Personal Auditory Displays

Federico Fontana, Davide Rocchesso and Laura Ottaviani
Proc. IEEE International Conference on Multimodal Interfaces (ICMI'02), pages 33–38, Pittsburgh, PA, October 2002.

A virtual resonating environment aiming at enhancing our perception of distance is proposed. This environment reproduces the acoustics inside a tube, thus conveying peculiar distance cues to the listener. The corresponding resonator has been prototyped using a wave-based numerical scheme called Waveguide Mesh, that gave the necessary versatility to the model during the design and parameterization of the listening environment. Psychophysical tests show that this virtual environment conveys robust distance cues.

G.1 Introduction

Humans use several senses simultaneously to explore and experience the environment. On the other hand, technological or human limitations often prevent computer-based systems from providing genuine multimodal displays. Fortunately, the redundancy of our sensory system can be exploited in order to choose, depending on cost and practical constraints, the display that is the most convenient for a given application.

Providing access to information by means of audio signals played through headphones or loudspeakers is very attractive, especially because they can elicit a high sense of engagement with inexpensive hardware peripherals. Namely, one may be tempted to transfer spatial information from the visual to the auditory channel, with the expected benefits of enlarging the perceptual bandwidth and lowering the load for the visual channel. However, we should bear in mind that vision and hearing play fundamental but different roles in human perception. In particular, space is not considered to be an "indispensable attribute" of perceived sounds [86].

In audio-visual displays, the effectiveness of communication can be maximized if the visual channel is mainly devoted to spatial (end environmental) information, while the auditory channel is mainly devoted to temporal (end event-based) information. However, there are several situations where the spatial attributes of sound becomes crucial:

1. In auditory warnings, where sounds are used to steer the visual attention;
2. Where it is important to perceive events produced by objects that are visually occluded or out of the visual angle;
3. For visually impaired users, where visual information is insufficient or absent.

Furthermore, if the "soundscape" of events being conveyed via the auditory channel is particularly rich, the spatial dislocation of sound sources certainly helps the tasks of separation and understanding of events, streams, and textures.

Much research has been dedicated to spatial auditory displays, with special emphasis on directional cues [124], but the literature on the perception and synthesis of the range of sources is quite limited [94, 95, 167].

What most psychoacoustic studies have found in this field is that we significantly underestimate the distance of sources farther than a couple of meters from the subject. The acoustic cues accounting for distance are mainly *monaural*:

1. *Intensity* plays a major role, especially with familiar sounds in open space. In the ideal case, intensity in the open space decreases by 6 dB for each doubling of the distance between source and listener [102].
2. *Direct-to-reverberant energy ratio* affects perception in closed spaces or reverberant outdoor environments. The reverberant energy comes from subsequent reflections of the direct sound, each of them having amplitude and time delay that vary with the characteristics of the enclosure, and with the source and listener's positions.
3. *Spectrum* conveys distance information as well, if the listener has enough familiarity with the original sound. In that case, spectral changes introduced in the direct signal by air loss and/or sound reflection over non-ideal surfaces can be detected by the listener, and hence reconducted to distance information [15].

Also, the existence of *binaural* cues has been demonstrated, these cues being particularly important in the case of nearby sources [23].

Among the monaural cues, the third could be exploited to display very large distances, because the spectral cues are relevant only for long paths. The first cue is not very useful in auditory displays and sonification, because it imposes restrictions to the listening level and it may lead to annoying soundscapes. The second monaural cue can be exploited to synthesize spatial auditory displays of virtual sources in the range of about ten meters. In order to do that, we have to use artificial reverberators to add reverberant energy to sounds.

The effects of the environment characteristics over sound are difficult to model and highly context-dependent. Moreover, the psychophysical process that maps the acoustics of a reverberant enclosure to the listener's impressions of that enclosure is still partially unknown [10]. For this reason, artificial reverberators are typically the result of a *perceptual* design approach [58], which has the fundamental advantage of leading to affordable architectures working in real-time, and has resulted in several state-of-the art realizations, providing high-quality rendering of reverberant environments [77]. Nevertheless, most of these realization do not deal with distance rendering of the sound source.

On the other hand, the *structural* design philosophy focuses on models whose properties have a direct counterpart in the structural properties that must be rendered, such as the geometry of an enclosure or the materials the wall surfaces are

made of. Thanks to that approach, their driving parameters translate into correspondent model behaviors directly. Unfortunately, structural models resulted to be either too resource-consuming, or, when simplified to accommodate the hardware requirements (i.e., the real-time constraint), excessively poor in the quality of the audio results.

The emerging *auditory display* field shifts the focus on the usability of the auditory interface rather than on the audio quality *per se*. For the purpose of enhancing the effectiveness of display, it is often useful to exaggerate some aspects of synthetic sounds. In spatial audio, this has led to the proposal of systems for supernormal auditory localization [43]. In this paper, we extend this concept to range localization of virtual sound sources. We use the structural approach to reverberation to design a virtual resonator that enhance our perception of distance. As a simple experience, consider a child playing inside one of those tubes that are found in kindergartens. If we listen to the child by staying at one edge of the tube, we have the feeling that she is located somewhere within the tube, but the apparent position turns out to be heavily affected by the acoustics of the tube. Using a virtual acoustic tool, we experimented with several tube sizes and configurations, until we found a virtual tube that seems to be effective for distance rendering. In a *personal* auditory display [95], where the user wears headphones and hears virtual as well as actual sound sources, these tubes will be oriented in space by means of conventional 3D audio techniques [124], so that the virtual sound sources may be thought to be embedded within virtual acoustic beams departing from the user's head.

This paper is organized as follows: first, we propose an ideal listening environment where "augmented" distance cues can be conveniently conveyed to a listener. Then, we turn this ideal environment into a structure-based sound processing model, that requires very simple and direct parameterization. Finally, the performance of the model is evaluated through a listening test where subjects use headphones, in order to assess the suitability of the proposed model for personal auditory display.

G.2 Acoustics inside a tube

The listening environment we will consider is the interior of a cavity having the aspect of a square-sectioned tube (size $9.5 \times 0.45 \times 0.45$ meters, see Figure G.1). The internal surfaces of the tube are modeled to exhibit natural absorption properties against the incident sound pressure waves. The surfaces located at the two far edges are modeled to behave as *total* absorbers [90]. The resonating properties of cavities having similar geometrical and absorbing characteristics (for example organ pipes) have been previously investigated by researchers in acoustics [48].

Although artificial, this listening context conveys sounds that acquire noticeable spatial cues during their path from the source to the listener. Given the peculiar geometry of the resonator, these cues mainly account for distance. The far surfaces have been set to be totally absorbent in order to avoid echoes originating from subsequent reflections of the wavefronts along the main direction of wave propagation. In fact, these echoes turned out to be ineffective for distance recog-

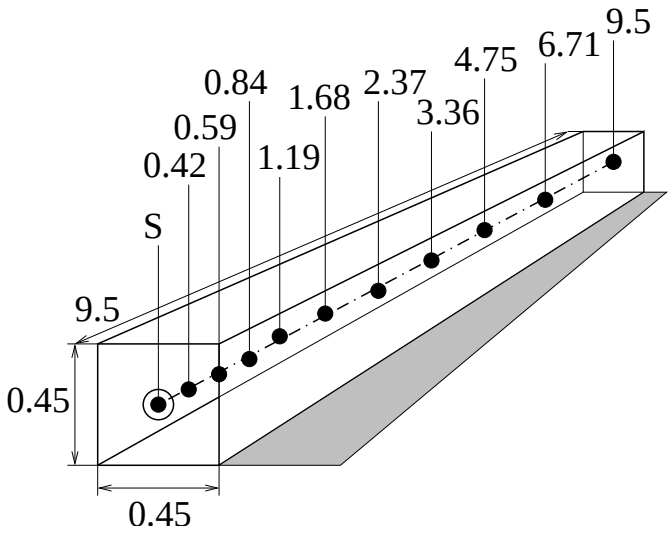

Fig. G.1. The listening environment. All sizes in meters.

nition in the range specified by the tube size, and they would also be annoying for the listener.

The tube size is the result of a prior investigation, where we experienced that thinner tubes caused excessive sound degradation, whereas thicker ones did not improve the basic functionality of providing distance cues, although resulting into computationally heavier models. Summarizing, we have designed a resonating environment that is structurally simple and computationally relatively light, meanwhile it deals with an interesting range of the physical quantity we aim at rendering.

In this environment, we put a sound source at one end of the tube (labeled with S in Figure G.1) along the main axis. Starting from the other end, we move a listening point along 10 positions $x_{10}, \ldots, x_1$ over the main axis, in such a way that, for each step, the source/listener distance is reduced by a factor $\sqrt{2}$. Finally the following set X of distances expressed in meters comes out, as shown also by Figure G.1:

$$X = \{x_i,\, i = 1, \ldots, 10\} = \{0.42, 0.59, 0.84, 1.19, \ldots, 4.75, 6.71, 9.5\} \qquad \text{(G.1)}$$

An obvious question arise prior to any investigation on distance rendering: why can't we render distance in an auditory display by simply changing the loudness of sounds as a function of their distance from the listener? The answer is twofold:

- Distance recognition by loudness is as more effective, as more familiar the sound source is. Conversely, a resonator, once become familiar to the listener, adds unique "footprints" to the sound emitted by the source, so that the listener has more chances to perform a recognition of distance also in the case of unfamiliar sounds.
- As introduced above, loudness in open spaces follows a 6 dB law for each doubling of the distance. This means that a wide dynamic range is required

for recreating interesting distance ranges in virtual simulations of open spaces. This requirement, apart from inherent technical complications coming from guaranteeing it at the hardware level, might conflict with the user's need of hearing other events (possibly loud) in the display. These events would easily mask farther virtual sound sources, especially in the case when the auditory display is designed to work with *open* headphones or *ear-phones* [87]. In other words, distance rendering is more effective if the loudness-to-distance law is not steep, due to both technical and perceptual reasons.

Summarizing, the proposed environment should lead to a robust rendering also with respect to unfamiliar sounds, and to a broad perceived range obtained by a compressed loudness-to-distance law.

G.3 Modeling the listening environment

The square tube has been modeled by means of finite-difference schemes. Since these schemes provide a discrete-space and time formulation of the fundamental partial differential equation accounting for three-dimensional wave propagation of pressure waves along an ideal medium [152], they clearly devise a structural approach to the problem of modeling a reverberant environment. Their ease of control is a key feature for this research, that came useful especially during the preliminary informal listening of several tubular environments differing in size, shape and absorption properties.

In particular, a *wave*-based formulation of the finite-difference scheme has been used, known as the Waveguide Mesh, that makes use of the wave decomposition of a pressure signal p into its wave components p^+ and p^- [158]. By adopting this formulation the spatial domain is discretized in space into equal cubic volume sections, and each of them is modeled as a lossless junction of ideal waveguides, scattering 6 input wave pressure signals coming from orthogonal directions, $p_1^+, \ldots, p_6^+$, into corresponding output waves, $p_1^-, \ldots, p_6^-$, respectively (see Figure G.2).

It can be shown that pressure waves travel along the Waveguide Mesh at a speed equal to

$$c_W \leq \frac{1}{\sqrt{3}} d_W F_s \tag{G.2}$$

where F_s is the sampling frequency, d_W is the waveguide length, and the symbol $<$ means that some spatial frequencies travel slower along the mesh. This effect is called *dispersion* [152], whose main effect is a detuning of high frequencies, which is not considered to be important for the application.

Assuming the velocity of sound in air equal to 343 m/s, and setting $F_s = 8$ kHz, we have from (G.2) that each waveguide is about 74.3 mm long. Thus, the mesh needs $127 \times 5 \times 5 = 3175$ scattering nodes to model our tube. Note that the sampling frequency has important effects on the computational requirement of the model. Clearly, our choice is oriented to efficiency rather than sound realism: reliable distance cues should be conveyed also using lower sampling frequencies.

The Waveguide Mesh has already been used in the simulation of reverberant enclosures [93]. In particular, it enables the use of *waveguide filters* at the mesh

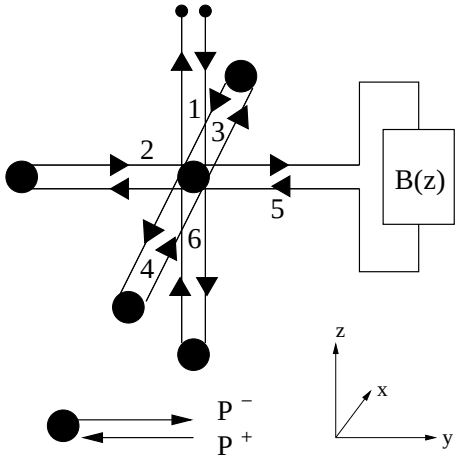

Fig. G.2. Particular of a volume section. The lossless scattering junction in the center is connected to other junctions via waveguides 2, 3, 4, and 6. Waveguide 1 leads to a totally absorbing section of wall. Waveguide 5 leads to a partially absorbing section of wall, modeled using a waveguide filter. The triangles filled with black represent oriented unit delays.

boundary, that model the reflection properties of the internal surfaces of the tube [75, 159].

Each waveguide branch falling beyond the boundary of the tube is terminated with a spring/damper system, that models the wall surface. This system is algebraically rearranged into a Waveguide filter, then discretized into a Waveguide Digital filter establishing a transfer function between pressure waves going out from, and incoming to the scattering junction: $B(z) = P_i^+ / P_i^-$. For example, it is $i = 5$ in Figure G.2.

Using the simple physical system seen above, the resulting model of the wall is made of 1st-order filters. Nevertheless, these filters model the properties of real absorbing walls with enough precision [90].

Considering that the surfaces at the two terminations of the tube have been set to be totally absorbing (this meaning that $p^+ \equiv 0$), the total number of boundary filters is $127 \times 5 \times 4 = 2540$.

G.4 Model performance and psychophysical evaluation

Ten stereophonic impulse responses have been acquired from the tube model along positions $x_1, \ldots, x_{10}$. The right channel accounts for acquisition points exactly standing on the main axis (refer to Figure G.1), whereas the left channel accounts for points displaced two junctions far from that axis, this corresponding to an interaural distance of about 15 cm. We then convolved these responses with a monophonic, short anechoic sample of a cowbell sound, and labeled the resulting 20 sounds according to the indexes and channels of the respective impulse responses: 1R,. . . ,10R and 1L,. . . ,10L.

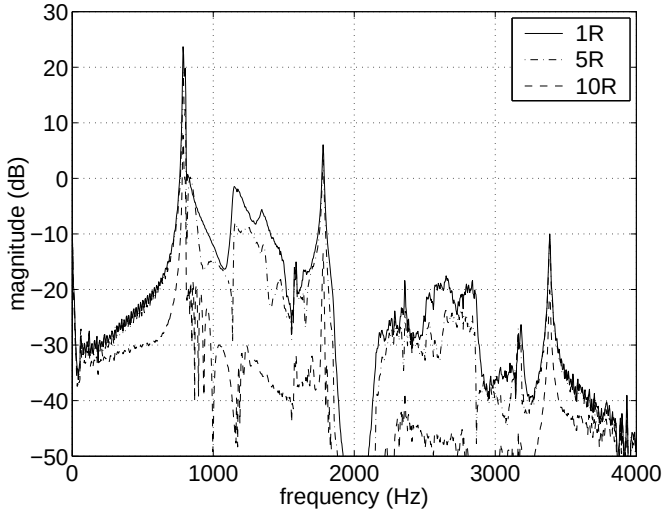

Fig. G.3. Magnitude spectra of signals 1R (solid), 5R (dash-dotted), 10R (dashed).

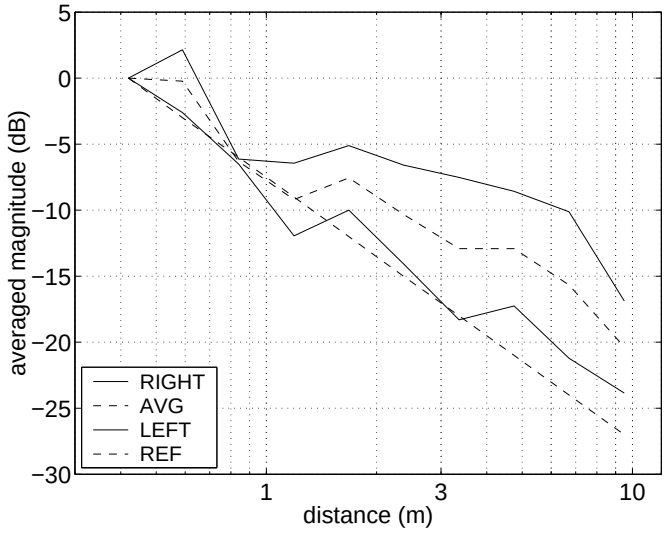

Fig. G.4. Averaged magnitudes of all acquired signals as a function of distance. Top (solid): right channel; middle (dashed): average between left and right channels; bottom (solid): left channel. Reference values in the ideal open-space case (bottom, dashed).

Measures conducted on those sounds are summarized in Figure G.3 and Figure G.4. Figure G.3 shows spectral differences existing between sounds auditioned nearby, mid-range and far from the sound source, for the right channel. The left channel exhibits similar differences. Figure G.4 shows how energies of signal s, defined by the value $10 \log \frac{1}{n} \sum_n [s(n)]^2$ (0 dB corresponding to the closest position),

vary with distance: These variations show that a dynamic range smaller than the 6 dB law is produced by the proposed method.

In particular, Figure G.4 shows that the right-channel magnitudes diverge from the left ones, as long as the range becomes greater than about 1 m. This divergence does not appear in reverberant sounds taken from real-world environments. This effect can be certainly reconducted to the peculiarity of the listening environment proposed here. Nevertheless, a careful analysis of the side-effects coming from using a coarse-grained realization of the Waveguide Mesh as a model of the listening environment should be carried out, to assess the precision of the plots depicted in Figure G.4.

We conducted also an experiment using the magnitude estimation method without modulus, to investigate how subjects scaled the perceived distance [151]. We asked 12 volunteers (4 females and 8 males, with age between 22 and 40), to estimate the distance from the 10 stereophonic sounds synthesized using the technique above. The experiment was performed in normal office background noise conditions. The setup involved a PC, a Creative SoundBlaster Live! audio card and Beyerdynamic DT 770 closed Hi-Fi headphones.

The set of sounds was repeated 3 times, for a total of 30 stimuli randomly distributed. Each subject had to associate the first stimulus with a distance, in meters, without prior training. Then, she evaluated the other stimuli proportionally to the first estimation. Since we did not set a modulus, the estimations define scales that depend on the individual listeners' judgments. These scales range from 0.2-8 (subject no. 8) to 1-30 (subject no. 5).

The three judgments given for each sound were then geometrically averaged for each subject, and the resulting values were used to calculate a mean average. Subtracting it from the individual averages, we adjusted the listeners' judgments to obtain a common logarithmic reference scaling [45].

In Figure G.5 the distance evaluations as functions of the source/listener distance are plotted for each subject, together with the corresponding linear functions obtained by linear regression. The average slope is 0.6093 (standard deviation 0.2062), while the average intercept is 0.4649 (standard deviation 0.2132).

In Figure G.6 the perceived distance averaged across values is plotted as function of the source/listener distance, together with the relative regression line ($r^2 = 0.7636$, $F(1,8) = 25.8385$, $F_{crit}(1,8) = 11.2586$, $\alpha = 0.01$). The r^2 coefficient is significant at $\alpha = 0.01$ and, therefore, the regression line fits well with the subjects' evaluations.

We observe that subjects overestimate the distance for sound sources that are closer to the listener. This overestimation tends to decrease as long as the distance increases. These results partially conflict with the conclusions of other studies, where distance perception was carefully assessed using sounds resulting from impulse responses captured from the real world [167]. Those studies have concluded that, under normal listening conditions, humans tend to overestimate the distance of nearby sources and, conversely, underestimate the range of sound sources that are far from the listener. The point of correct estimation occurs at about 1-1.5 meters.

One explanation for the results obtained here can be found in the exaggeration of the reverberant energy that is produced by the peculiar listening environment

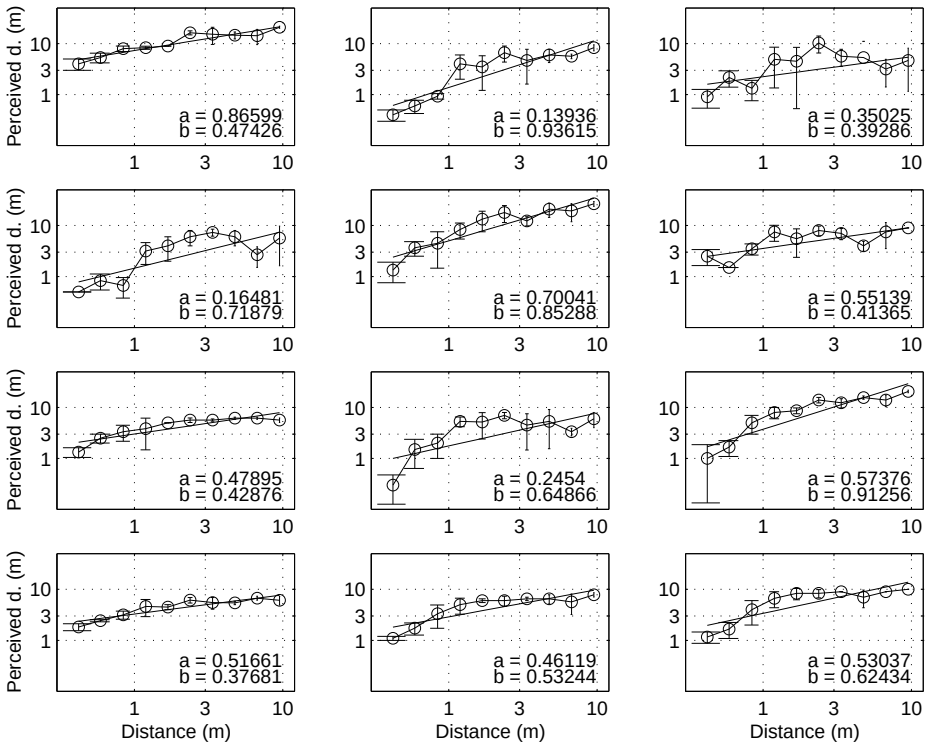

Fig. G.5. Individual distance evaluations together with individual linear regression lines. a: intercept. b: slope.

we adopted, which seems to offset the listeners' evaluation toward higher ranges. Indeed, this exaggeration may help to set up an auditory display where sounds in the far-field must be reliably reproduced.

An important remark is that in general subjects did not complained about the lack of externalization in the stimuli, even though the experiment was conducted without any binaural spatialization tool [124], and using regular headphones rather than special equipment [87]. Only one subject (a graduating student who is not novice in the field) attributed a null distance to the stereophonic sound (1L,1R) in two of the three judgments, later reporting on some in-head localization of those two samples. Therefore, we conclude that the impression of distance can be rendered quite effectively and economically just by proper design of a reverberating environment.

G.5 Conclusion

A virtual listening environment capable of sonifying sources located in the far-field has been presented, along with a versatile way to model it. Listening tests show

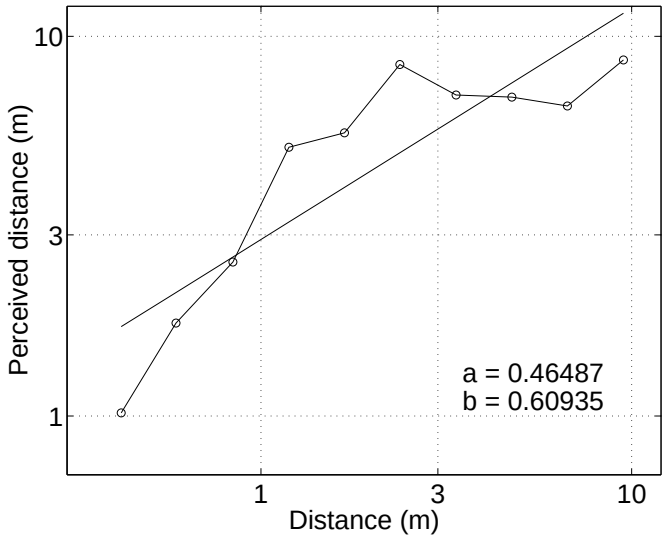

Fig. G.6. Average distance evaluation together with linear regression line. a: intercept. b: slope.

that it actually conveys exaggerated range cues, nevertheless the model can be easily re-parameterized to account for different psychophysical scales.

G.6 Acknowledgments

Federico Beghini and Fabio Deboni have collaborated in the modeling of preliminary virtual listening environments. The Authors are also grateful to the volunteers who participated to the listening test.

Partial support has been given by the Project SOb - The Sounding Object (http://www.soundobject.org), as part of the European Commission's Future and Emergent Technologies collaborative R&D programme.

H

A Digital Bandpass/Bandstop Complementary Equalization Filter with Independent Tuning Characteristics

Federico Fontana and Matti Karjalainen
IEEE Signal Processing Letters.
- IN PRESS -

A discrete-time realization of first-order (shelving) and second-order equalization filters is developed, providing bandpass/bandstop magnitude-complementary transfer functions. The bandpass transfer function is turned into a complementary one, and vice versa, switching between two structures that share a common allpass section containing the state variables of the equalizer. The present realization is an alternative to existing solutions, particularly in applications where the equalization parameters are dynamically varied.

H.1 Introduction

Linear equalization is a well-known processing technique, used when a musical signal needs improvements in "presence" or exhibits artifacts which can be corrected through a manipulation of its spectrum. In the digital domain graphic equalization can be efficiently performed using first-order and second-order *tunable equalization filters* [119]. In such structures, the low-frequency and high-frequency gains (first-order equalization filters) and the center-frequency gain (second-order equalization filters) are determined by the value of one multiplying coefficient. Other tuning parameters, i.e., the cutoff frequency of first-order filters, and the selectivity and center frequency of second-order filters, are embedded in the coefficients of an allpass block, whose freedom of realization is another valuable feature of these structures [99].

Although such filters are minimum phase, their gain responses cannot be complemented (more specifically, their transfer functions cannot be reciprocated) just by varying the gain coefficient. Many solutions have been proposed to overcome this problem. In some cases proper functions have been designed, mapping the gain parameter over the filter coefficients of the equalizer [20, 38]. In other cases the allpass section of the tunable equalization filter has been preserved during the calculation of the inverse transfer function, yielding an efficient realization of the complementary equalizer [170].

In this paper a different approach is presented. Given a minimum-phase transfer function $H(z) = 1 + A(z)$, where $A(z)$ is an allpass filter where pure unit delays cannot be factored out, the inverse $1/H(z)$ can be realized using a procedure which preserves most of the structural properties of the original structure, in particular the allpass block [69].

With this procedure we can compute the inverse transfer function of a bandstop tunable equalization filter. In this way we obtain a complementary bandpass response avoiding the design of a new, independent filter. The complete bandpass/bandstop equalization filter is made of two different structures which are alternatively used anytime the response switches from bandpass to bandstop and vice versa. In spite of this, the allpass section is left untouched by such commutation.

Compared with the Regalia-Mitra filter [119], a real-time implementation of the proposed system demands more computation cycles on a digital signal processor. However, the lookup of filter coefficients is simplified. Its performance becomes valuable in application cases involving dynamic variations of the gain parameter.

H.2 Synthesis of the Bandpass Transfer Function

The transfer function of a bandstop tunable digital equalizer, H_{BS}, can be put in the form

$$H_{\mathrm{BS}}(z) = \frac{1+K}{2}\left\{1 + \frac{1-K}{1+K}\,A(z)\right\}\,,\ 0 < K < 1 \tag{H.1}$$

where $A(z)$ is an allpass filter [119]. When $A(z)$ is a first-order allpass then we have a low-frequency (LF) shelving filter. When $A(z)$ is a second-order allpass then we have a second-order equalizer. An LF shelving filter is turned into a high-frequency (HF) one by changing the sign of $1 - K$ in (H.1). The value of K determines the edge-frequency cut in the case of shelving filters, and the notch-frequency level in the case of second-order equalizers. In both cases, H_{BS} is minimum-phase.

Let $H_{\mathrm{BP}}(z)$ be the inverse of $H_{\mathrm{BS}}(z)$. An inversion of the bandstop transfer function aiming at preserving the allpass block induces the *delay-free loop problem* [98, §6.1.3]. In fact, as long as a new value of the input signal x is acquired at time step n, we cannot explicitly calculate the bandpass filter output y making use of the output l from the allpass:

$$y[n] = \frac{2}{1+K}\,x[n] - \frac{1-K}{1+K}\,l[n] \tag{H.2}$$

since the computation of $l[n]$ requires the knowledge of $y[n]$.

Nevertheless, we can look at $l[n]$ as a linear combination of two components, i.e., $l[n] = \alpha y[n] + l_0[n]$: the first component, accounting for the delay-free part of the allpass response, contains the coefficient α, that sets the cutoff frequency and the selectivity of first- and second-order equalizers, respectively [119]; l_0 is the output from the allpass filter fed with zero. By substituting this formula in (H.2), it can be seen that $y[n]$ can be computed at each time step using the following procedure [69]:

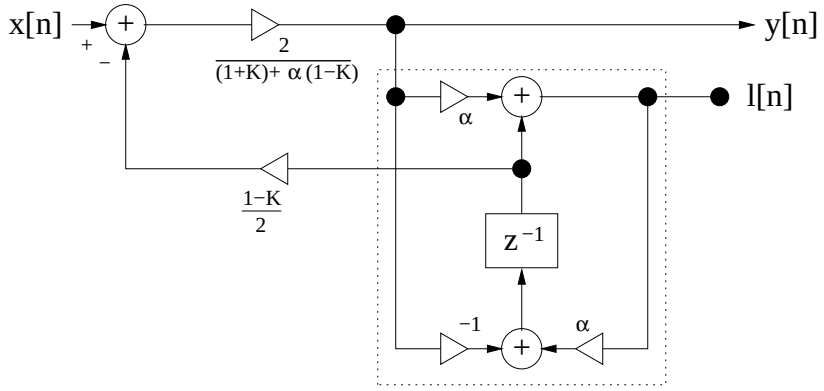

Fig. H.1. Switch-free structure for the computation of the bandpass transfer function of a shelving filter. The allpass section, realized in transposed canonic form, is located inside the rectangle in dashed line. The terms $1 - K$ must be changed into $K - 1$ when performing HF shelving.

1. $l_0[n]$ is computed feeding the allpass filter with zero;
2. $y[n]$ is calculated by

$$y[n] = \frac{2}{(1+K) + \alpha(1-K)} \left\{ x[n] - \frac{1-K}{2} \, l_0[n] \right\} \tag{H.3}$$

3. the allpass filter is fed with $y[n]$ to update its state variables.

This procedure can be computed providing the filter structure with switches that, during each time step, alternatively feed the allpass block with zero or $y[n]$, respectively implementing steps 1 and 3 of the procedure. The use of switches allows to design the allpass independently of the rest of the structure. Otherwise, switches can be avoided if we reconsider the allpass output decomposition as $l_0[n] = l[n] - \alpha y[n]$ [70]. As an example, Figure H.1 shows a switch-free structure that embeds an allpass filter in transposed canonic form, realizing a shelving filter. This example can be immediately extended to second-order realizations, substituting the first-order allpass with a second-order allpass in the same form.

A structure which includes switches for computing the above procedure is included as part of the system depicted in Figure H.2, inside the rectangle in dashed line. In that structure, a hold block (labeled with H) retains the output that is calculated when the switches are in position I (steps 1 and 2 of the procedure), then feeds the allpass block when the switches are in position II (step 3).

The whole discussion still holds if the range of K is changed into $1 < K < \infty$, i.e., if the roles of H_{BS} and H_{BP} are swapped. Such a choice is supported by the definition of cutoff frequency and filter selectivity [119] governing the value of α. However, the former assumption results in filter coefficients whose range better accommodates fixed-point number representations.

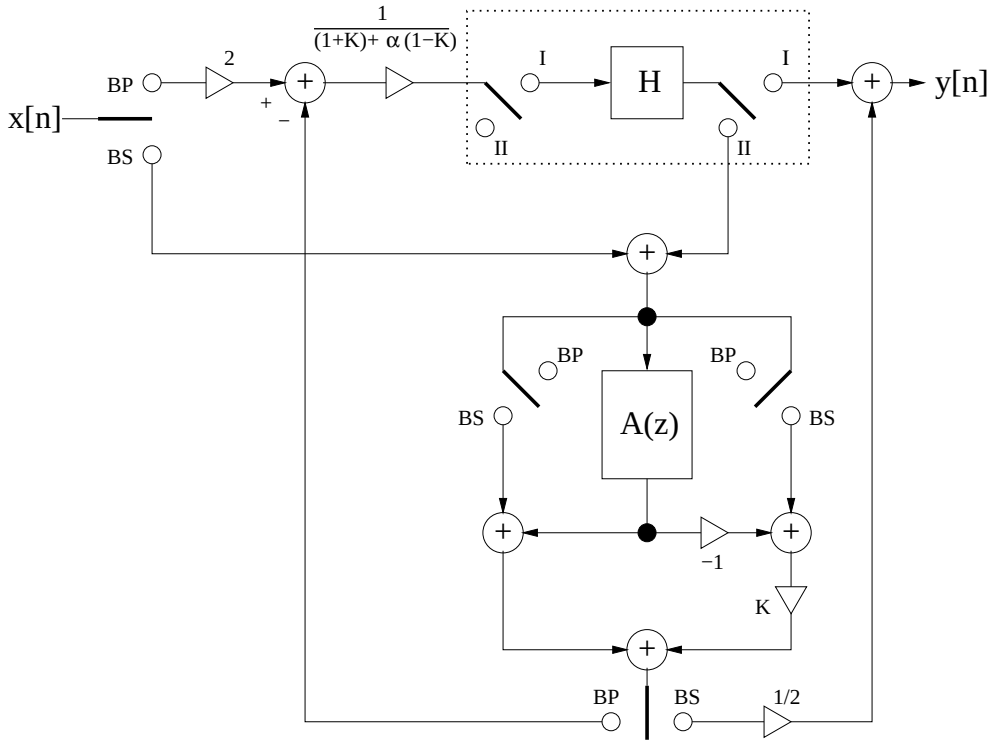

Fig. H.2. Bandpass/bandstop equalizer. When switches are set to position BS a bandstop transfer function is provided. When switches are set to position BP a complementary (bandpass) transfer function is provided. A switching structure (located inside the rectangle in dashed line) containing a hold block H is employed for calculating the inverse transfer function. HF shelving is realized by swapping the branch containing the multiplier by K with the one terminating at the adder common to both of them.

H.3 Implementation of the Equalizer

A system realizing both the bandpass and bandstop transfer function is given in Figure H.2. When the switches are in the position labeled with BS, the system implements a bandstop equalization filter. When they switch to position labeled with BP, the system becomes a complementary (bandpass) filter. In principle, transitions between the two structures do not introduce transients in the output signal and the internal state. In fact they happen when $K = 1$: under this condition the position of the switches is insignificant.

Figure H.3 shows (in dashed line) plots of six gain responses from the original Regalia-Mitra LF shelving filter for gains equal to -12, -8, -4, 4, 8 and 12 dB, respectively, together with plots (solid line) of the responses from the proposed system having the same gains of the Regalia-Mitra filter when working in bandstop configuration. In all cases $\alpha = 0.8668$. Figure H.4 shows similar gain responses for the second-order equalizer (dashed line), centered at a normalized frequency equal to 0.01, together with gain responses from the proposed system (solid line),

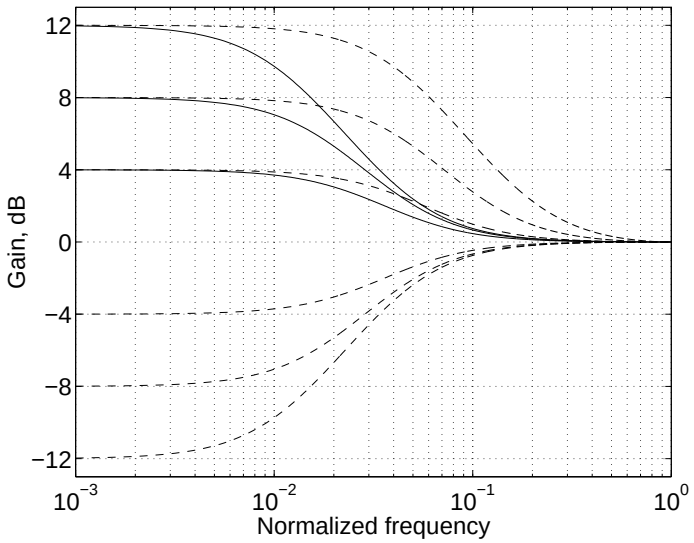

Fig. H.3. Gain responses from the Regalia-Mitra LF shelving filter for gains equal to -12, -8, -4, 4, 8 and 12 dB (dashed line). Gain responses from the proposed system for gains equal to -12, -8, -4 dB (solid line). In both cases $\alpha = 0.8668$.

complementing the bandstop Regalia-Mitra equalizer ($\alpha = 0.9844$). Moreover, six responses coming from a second-order equalizer that has been designed following the Bristow-Johnson approach [20] are also plotted, for the same gains and setting the center frequency to 0.21.

In the proposed system additional computations are needed to calculate $l_0[n]$ and $y[n]$. In particular, they reduce to one sum plus one multiply to compute $y[n]$ and one multiply followed by one sum to compute $l_0[n]$, if the allpass section is realized with a lattice structure. One table lookup is required to load the value $G_{\mathrm{BP}} = 1/\{(1+K) + \alpha(1-K)\}$ each time a variation in the parameter K occurs. A multiplication by two must be applied to the input signal when the switches are set to position BP. Conversely, the output signal is divided by two when the system is in bandstop configuration.

Such result seems valuable if we compare the system in Figure H.2 with Regalia-Mitra filters, in front of variations of the gain parameter. With those filters we can require magnitude-complementary bandpass/bandstop transfer functions. If we choose to preserve the independence of gain tuning in the bandstop filter, then a complementary bandpass response can be obtained by tuning the parameter α_{BP}, that will be used in place of α in bandpass configuration, according to the value of $K > 1$:

$$\alpha_{\mathrm{BP}} = \frac{1 - K^{-1} + \alpha(1 + K^{-1})}{1 + K^{-1} + \alpha(1 - K^{-1})} \tag{H.4}$$

(as usual, when performing HF shelving we have to change $1 - K^{-1}$ into $K^{-1} - 1$).

Hence, a complementary tunable equalization filter requires looking up one table stored with coefficients $K < 1$ when providing the bandstop response, and

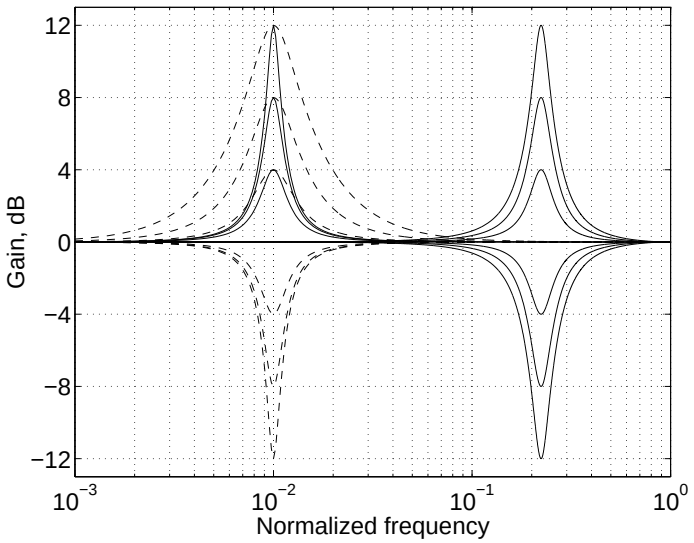

Fig. H.4. Gain responses from the Regalia-Mitra second-order equalizer for gains equal to -12, -8, -4, 4, 8 and 12 dB (left, dashed line). Gain responses from the proposed system for gains equal to -12, -8, -4 dB (left, solid line). In both cases $\alpha = 0.9844$. Normalized center frequency set to 0.01. Gain responses from the Bristow-Johnson equalizer for gains equal to -12, -8, -4, 4, 8 and 12 dB (right, solid line). Normalized center frequency set to 0.21.

two tables stored with values $K > 1$ and α_{BP}, respectively, when providing the bandpass transfer function[1].

The proposed system equals a tunable equalization filter when providing a bandstop response. Since the bandpass transfer function is obtained through a switch, it needs neither a table stored with $K > 1$, nor a table accounting for α_{BP}. However, it needs a table stored with G_{BP}, that is looked up according to the value $K < 1$. For this reason, the coefficients in the system have magnitudes that are inherently smaller than unity.

In summary, a comparison between the proposed system and complementary Regalia-Mitra equalizers, made with respect to variations of the gain parameter, shows that the former solution prevents from storing gain coefficients $K > 1$, needed for bandpass filtering. This advantage is paid computationally, since two more sums and two more multiplies must be executed during each time step in bandpass configuration.

H.4 Summary

A digital equalizer has been presented. It integrates previously known tunable equalization filters which are used here to provide the bandstop transfer function. The system switches to a different structure providing a complementary (bandpass) transfer function, meanwhile it preserves the internal state and most of the

[1] such values are properly scaled when they are represented in fixed-point arithmetic.

structural properties of those filters. This property results in less memory consumption and simplified mapping of the equalization parameters over the filter coefficients.

Acknowledgments

Prof. Mark Kahrs and Prof. Vesa Välimäki gave many suggestions for improving the paper. This work has been partially funded by the "Sound Source Modeling" Project of the Academy of Finland.

I

Characterization, modelling and equalization of headphones

Federico Fontana and Mark Kahrs
Journal of the Virtual Reality Society.
- SUBMITTED FOR REVIEW -

The reproduction of sounds in individual auditory displays relies now on methods that are capable of localizing a sound source in a three-dimensional virtual acoustic scenario. The existence of such methods has created a novel interest in headphones as means for conveying information to the user through the auditory channel. Although their optimal design is a matter still under discussion, headphones represent the cheapest way to reproduce a virtual acoustic scenario to individuals.

In spite of many promises coming from the market, it is true that a precise characterization of several types of headphones can be given. This characterization starts from psychoacoustic arguments, and motivates a particular design using proper electro-acoustic models that, finally, give reason of the performance and the limits of a headphone model. Nevertheless, traditional Hi-Fi equipment is not optimally suited for those localization methods, that usually require binaural listening conditions: just a slightly incorrect positioning of the headphone cushion over the ear, for example, can alter such conditions producing significant localization distortion.

Most of these unsolved questions have recently inspired new headphone designs, that have dramatically simplified the conditions for binaural listening. These headphones will have an important role in the VR installations of the years to come.

I.1 Introduction

Loudspeaker design relies on basic and well-known desiderata. The most important aspect is the *transparency* of the loudspeaker system with respect to the sound to be reproduced and the impact on the listening environment. On the contrary, the optimal design of a headphone is not completely clear. This uncertainty has let manufacturers follow several philosophies, requiring both the reproduction of the free-field response and the diffuse field response. The question of the listener's

comfort has brought different solutions as well, such as the circumaural or supra-aural design.

Some recent methods, especially developed for the three-dimensional localization of a sound source in the reproduction of virtual acoustic scenarios [8] [4], have raised the question about which headphone is best for working together with these methods, i.e., which headphone design matches well with the ideal of transparency.

The wide choice of equipment available for Hi-Fi ranges from in-ear headphones for portable CD players to exotic and very expensive models [142]. Some manufacturers claim to offer audio equipment capable of reproducing virtual scenarios, especially designed to work together with normal Hi-Fi headphones [40].

In this paper we try to characterize headphones and headphone listening. First, important psychoacoustic aspects related to headphone listening are addressed. Then, headphones are modelled as equivalent electrical networks, so that the advantages and disadvantages of common types of headphones are exposed. Finally, a review on the research conducted in the field of binaural listening is presented, with particular attention on recent results, and together with a look at some equipment especially designed for binaural reproduction.

I.2 Psychoacoustical aspects of headphone listening

In principle, there is no difference between natural hearing and headphone listening, if the sound pressure at the two ear canal entrance positions provided by the headphones equals the pressure of the natural sound. In most cases this ideal condition is not achieved, so that the former falls into a non-ideal, quite unpredictable experience influenced by aspects such as the coding of the audio material, quality of the equipment, type and positioning of the headphones, and more.

Despite this, the audible artifacts are almost always limited to unnatural colouring of the sound, lack of precision in source positioning, and in-head localization (IHL). All these artifacts can be motivated using arguments from psychoacoustics.

I.2.1 Lateralization

Lateralization is the discrimination of sidedness in localization. It is known to depend mainly on the Interaural Time Difference (ITD) between the ears for low frequency signal components, and on the Interaural Level Difference (ILD) for high frequency components [15].

When stereophonic material is played in an anechoic listening environment using loudspeakers, intensity differences at low frequencies are perceived as time differences. In fact, the head shadow gives no attenuation at low frequencies, so that the superposition of pressure vectors at the ear entrances results into an ITD (Figure I.1) [15] [124]. This consideration is sufficient to conclude that stereophonic material intended for loudspeaker listening cannot give precise lateralization cues at low frequency if heard using headphones.

Conversely, at higher frequencies the head shadow attenuates sounds at the far ear, and the ITD becomes insignificant for source localization. In this case

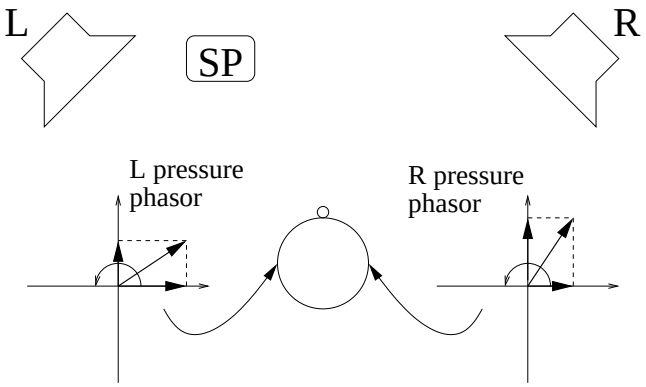

Fig. I.1. ITD by stereophonic panning using loudspeakers. SP is the position of the virtual source so that the left loudspeaker plays louder. Panning results from a phase shift between the two pressures measured at the ear-entrance positions.

panning provides lateralization cues both in the case of loudspeaker and headphone listening. With complex signals, lateralization depends also on the ITD of signal envelopes which are detectable in the whole audio range. One example of this is the ITD of a signal onset.

Satisfactory lateralization cues can be created from a monophonic signal, using the structural model described in [21]. Modelling a simplified (spherical) head, two signals are produced having phase shift and amplitude difference at high frequency, corresponding to ITD and ILD amounts needed to lateralize the sound source of a given azimuth. With a careful selection of the model parameters, precise lateralization cues have been obtained using monophonic piano samples as sound sources. Clearly, a model like this is useful only if monophonic material is available, and in any case it cannot provide lateralization of multiple sources unless they are separately recorded.

I.2.2 Monaural cues

As the position of the sound source approaches the median plane, monaural (spectral) cues become more prominent. Using these cues, a precise localization of the sound source becomes more problematic. Even if still related to binaural listening, ILD cues are known to result in unstable information, causing the listener to locate the sound source on the surface of a cone whose geometrical parameters are defined by the ITD, the so called cone of confusion (see Figure I.2) [15].

This cone degenerates into a disk in the median plane, and confusion results in front-back inversion and elevation error. Sources located on the median plane can be localized only by spectral cues embodied in the signal measured at the ear canal entrance. Experiments have shown that specific bands account for different directions (without precise localization): this result is explained by the directional-band model of median-plane localization [15]. Monaural cues apply only in the case that a preliminary learning process of the sound has been conducted (consciously or subconsciously) by the listener, in fact they can explain the localization process of

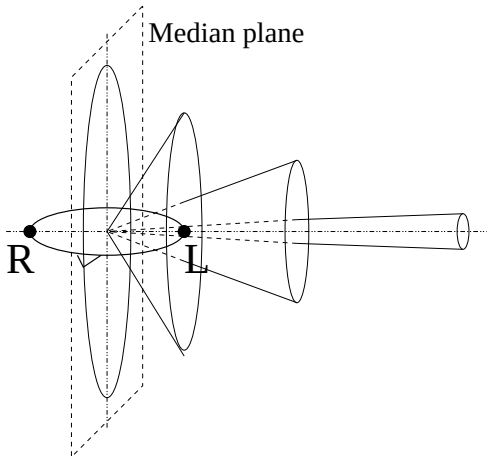

Fig. I.2. Cones of confusion for lateralization cues given by various ITDs (far ear on the right side).

familiar sounds only. More in general, head movement becomes an essential mechanism in the perception of elevation, otherwise the spectral cues of the Head-Related Transfer Functions (HRTF) accounting for elevation are often barely meaningful.

For all these reasons, elevation cues are difficult to reproduce using normal equipment. Experiments on perception of elevation far from the median plane have been conducted using normal headphones [8]. One remarkable problem lies in the fact that headphone listening cannot account for dynamic cues, unless providing the installation with a processing system controlled in real-time by head tracking. Nevertheless, important interference in the experiments comes from the listener's awareness of wearing headphones.

Finally, an intrinsic difficulty in reproducing frontal localization would suggest that the mechanisms of perceiving elevation are still not completely understood. This difficulty comes also from the fact that binaural cues also occur during the recognition of elevation, due to pinna disparity. The importance of these cues has been demonstrated by listening tests using subjective HRTFs individually measured at each ear: listeners, asked to rate the quality of audio materials in terms of elevation and IHL perception, gave the best score to recordings played using individually measured HRTFs [141].

I.2.3 Theile's association model

The association model of Theile (Figure I.3) gives a convincing interpretation of the unnatural colouring of sounds listened using headphones [154]. In this model it is suggested that distortions (denoted with M) introduced in the binaural signal by the outer ear are removed by a canceling stage, M^{-1}, if they are recognized *for what they are*, i.e. if additional information coming from the external world suggests a correct positioning of the sound source according to the acoustic information. Otherwise the stage M^{-1} is omitted, thus leaving unnatural color on the sound. Of

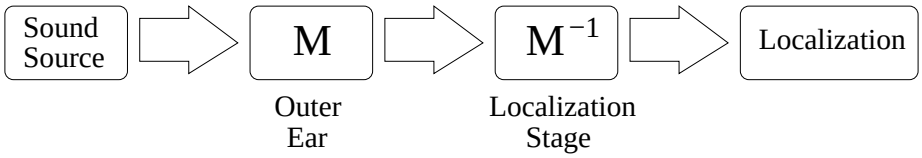

Fig. I.3. Theile's association model. Stage labeled with M accounts for sound distortions introduced by the outer ear.

course, this model is not useful in establishing localization models for headphone listening.

I.2.4 Other effects

IHL occurs especially because of the lost precedence effect [15], which otherwise plays a fundamental role in loudspeaker reproduction. The suppression of this effect emphasizes the importance of a careful reverberation as perhaps the most useful strategy to externalize sounds using normal headphones. In particular, uncorrelated binaural signals obtained by proper processing of a monophonic signal have been shown to give a certain degree of out-of-head localization (OHL) during headphone listening tests [82].

In the open ear case, the ear entrance can be seen from the inner ear as a low-radiation piston, i.e. a resistance R_{OE} in parallel with an inductance L_{OE} if the electrical equivalent is considered. Occluding it transforms the ear canal into a pressure chamber, turning the radiation impedance into a volume compliance or, from the equivalent electrical network point of view, substituting the parallel $R_{OE} \| L_{OE}$ with a capacitance C_{CE} (see Figure I.4). This change has the noticeable effect of changing the transfer characteristic of the ear canal, producing a boost of the frequencies below approximately 2 kHz. If a carefully made headphone tries to account for this, the occlusion effect allows bone conduction to be heard, i.e. the skull vibrations resonate in the pressure chamber. Also, occlusion can result in the transmission to the eardrum of ultrasounds up to 100 kHz, which are heard in the form of a constant pitch around the upper frequency limit of the hearing mechanism [11].

I.2.5 Headphones vs. loudspeakers

Most of the differences between headphone and loudspeaker listening arise from the questions discussed above. A well-known question lies in the so called "missing 6 dB" below 300 Hz, whose explanation is quite clear for low sound pressures: sounds are masked by physiological noises amplified by the occlusion effect. When the sound pressure is high (typically above 40 dB), the explanation is more difficult. In this case, the effect seems to depend on mechanical vibrations created by the loudspeakers, the psychological "larger acoustic size" of the loudspeaker, and possible distortions coming from amplification [131].

Under normal conditions, i.e., in a reverberant room, phase is inaudible with loudspeaker listening. Subtle effects due to phase distortion can be heard using

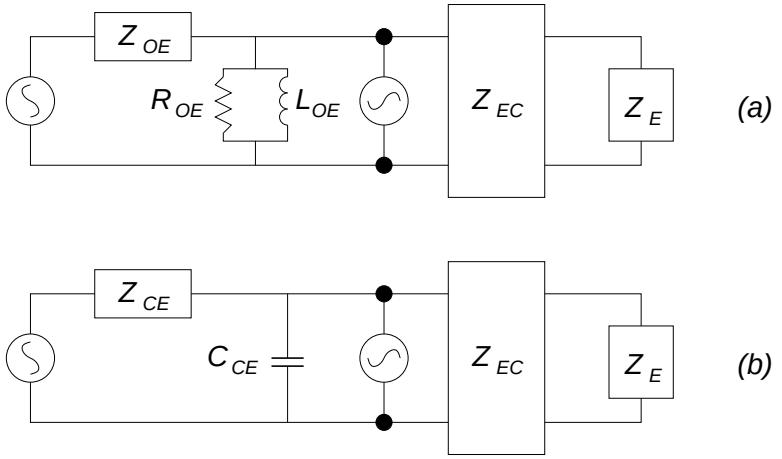

Fig. I.4. Occlusion effect. Equivalent circuit of the (a) open and (b) occluded ear canal seen by the inner ear. The current source accounts for signals coming to the ear-canal through bone conduction. The $R_{OE} \| L_{OE}$ parallel becomes a capacitance C_{CE} when the ear is occluded. Impedance value at the ear canal entrance Z_{OE} changes in Z_{CE}. The equivalent ear canal and eardrum impedances, Z_{EC} and Z_E respectively, are left unchanged.

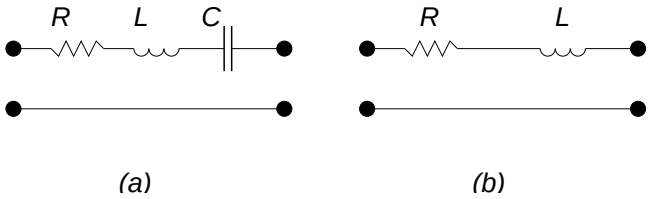

Fig. I.5. Some equivalent circuits: (a) piston membrane; (b) hole or slit.

headphones. This suggests that non-minimum phase stages in the reproduction chain should be designed with particular care, when headphones are adopted as voltage-to-pressure transducers at the end of the audio reproduction chain.

I.3 Types and modelling of headphones

The design of a high-quality headphone is a matter of experience and taste. Despite this, its main building blocks can be modelled as acoustic elements yielding pressure-flow relations in lumped systems: volume compliances, radiating pistons, piston membranes, holes, slits. Each one of these elements has a direct counterpart in the world of electrical networks. The electrical equivalent circuits of the volume compliance and the radiating piston have already been shown in Figure I.4. Figure I.5 shows the remaining ones. The lumped approach is valid up to a few kilohertz. Above typically 2 kHz, non-uniform distributions of pressure must be taken into account, and the electrical equivalent gives only qualitative information.

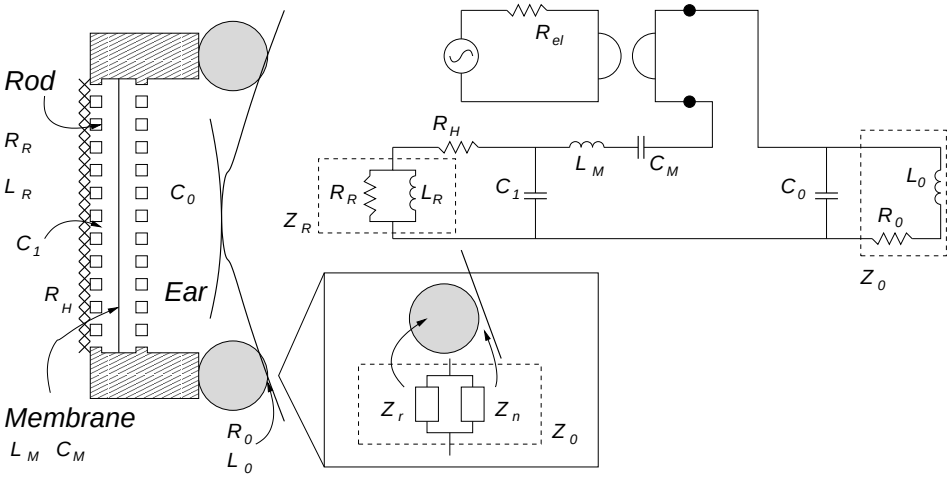

Fig. I.6. Schematized isodynamic headphone (left) and its electroacoustic model (right). Particular of the electrical equivalent of the cushion in contact with the head

The acoustic system then must be coupled with the electrical circuitry, to form an electroacoustic equivalent circuit, again represented by an electrical network. The electroacoustic coupling is realized by means of acoustic transformers and gyrators, which transduce pressure-flow into force-velocity relations, coping with the voltage-current formulation of the electrical network.

In this section, most of the common types of headphones are characterized as electroacoustic systems, to analyze their pros and cons related to their design philosophy. A more detailed treatment of the arguments presented in sections I.3.1 and I.3.2, and a comprehensive explanation of the electroacoustic equivalent circuits, can be found in [113], if not otherwise specified.

I.3.1 Circumaural, supra-aural, closed, open, headphones

The frequency response of a headphone strongly depends on how the transducer is coupled with the ear. This coupling is often quite uncertain, as it changes according to the position of the headphone over the head.

With the *circumaural* design, the earpiece is well sealed inside the pressure chamber existing between the cushion and the head. This kind of headphone provides good insulation from external noise. The intrinsic stability of the chamber's parameters results in reliable electroacoustic models, however the relatively large volume compliance increases the sensitivity at low frequency with respect to variations of this volume, and reduces the range where the lumped approach is valid. Circumaural headphones are always *closed*.

Supra-aural headphones in principle have less reproducible frequency responses due to unpredictable leaks between the cushion and the earpiece. Recently, these headphones have improved in performance, and they are often chosen even for high quality sound reproduction in low-noise environments since they have the advantage of lightness. The precision of the low-frequency response of supra-aural

headphones is increased by dividing the foam of the cushion into a part providing the best contact with the head, thus minimizing unpredictable leaks, and another part, usually rigid and porous, providing a rather high, stable and predictable leak. In this way the total leak Z_0 (see the particular of the cushion in Figure I.6) of the acoustic chamber is represented by a parallel of two impedances, Z_r and Z_n, where the non-reproducible leak Z_n has a minor effect in the equivalent impedance $Z_0 = Z_r \| Z_n$. The reduced volume of the pressure chamber in the supra-aural design extends the validity of the lumped approach since standing waves appear at higher frequencies.

Supra-aural headphones can be provided with holes and/or membranes located on the external shell. In this case they are called *open*. Holes are used to gain further predictability of Z_0. In order to counteract the contribution of the reactive component in the leak (which increases with frequency (see Figure I.5)), membranes can be properly placed instead of holes. Integrated open headphones with passive membranes often exhibit the most stable low-frequency response. Moreover, experiments show that more OHL localization is experienced using open headphones [71]. Even if some cross-talk occurs with the open design, the reasons for the increased OHL seem to be due to low sound reflection inside the pressure chamber, allowing a more undisturbed pinna activity, and a non-negligible perception of the external noise.

I.3.2 Isodynamic, dynamic, electrostatic transducers

Today, almost all high-quality headphones are provided only with isodynamic, dynamic, or electrostatic transducers, sometimes with more than one type in multiway systems.

The *isodynamic* transducer consists of a large-area, low-mass surface where a conductor driven by the audio signal forms a long track covering the whole surface. A magnetic field perpendicular to this track is created by small magnetized rods, oriented in such a way that they repel each other.

The electroacoustic model accounting for circuitry, transducer and ear-coupling is given in Figure I.6 (right), together with a general scheme of the isodynamic headphone (left). The network emphasizes the presence of a resonant circuit mainly governed by the membrane parameters, C_M and L_M, and the volume compliance given by the pressure chamber, C_0, typically producing a strong peak at around 4 kHz. This peak is damped by the resistive component R_R of leak Z_R and, if necessary, by a porous cover over the transducer providing a further resistance R_H. With this design, the leaks between cushion and head ($Z_0 = R_0 + j\omega L_0$) and the small volume compliances (C_1) between the rods have far less impact than the main leak $Z_R + R_H$ and the volume compliance C_0. The large moving area, represented by the isodynamic transducer, over the relatively small pressure chamber allows extension of the lumped approach until approximately 5 kHz. Within this band, very flat responses can be obtained by a careful tuning of the parameters.

Dynamic headphones take advantage of the powerful pressure produced by the moving-coil transducer (approximately ten times the pressure generated by the isodynamic transducer). Once again, an electroacoustic model can be developed

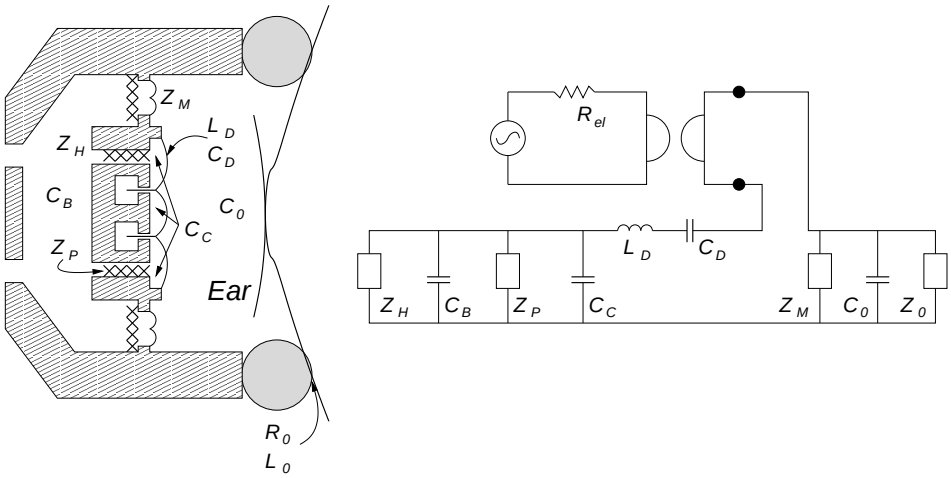

Fig. I.7. A general dynamic headphone (left) and its reduced electroacoustic model (right).

(Figure I.7, right)[1] once the most important elements in this design have been taken into account (Figure I.7, left). Again, the main resonance is governed by the transducer parameters, C_D and L_D, and the volume compliance, C_0. Resonance typically occurs at 600 Hz. The stronger sound pressure requires a comparable damping, given by leak Z_0, resistance R_P within the component Z_P accounting for porous material put behind the transducer, and the resistive component of the impedance Z_M given by passive membranes elements coupling with the pressure chamber. Non-negligible components in the computation of the low-frequency response are given by the volume compliances existing behind the transducer, which are summarized by the capacitance C_C, the volume behind the pressure chamber (C_B), and the holes necessary for the transducer membrane to work correctly (Z_H, refer also to Figure I.5).

The low-frequency response of the dynamic headphone exhibits less flatness than the isodynamic case, as damping is usually more complicated to design. The reliability of this analysis is also limited to a frequency of about 3 kHz. The moving coil transducer is much more efficient than the isodynamic one, as discussed earlier.

Electrostatic transducers (Figure I.8) are made by polarizing (typically 200 V) a membrane and then driving it with an electrostatic field modulated by the audio signal[2]. Polarization must be kept constant. This is often produced using the power coming from the audio signal itself, after rectification, and introducing a large resistance in series with the membrane. A drawback of this design lies in the slow re-polarization of the membrane when, during high signal levels, it touches the electrodes which form the electrostatic field, hence discharging the capacitor. Thus, the large resistance is often spread over the membrane, thereby limiting its discharge at the contact point.

[1] the model can be reduced to the network shown here only after a first, more rigorous scheme has been carefully studied and simplified

[2] typically using a push-pull circuit, to minimize distortion

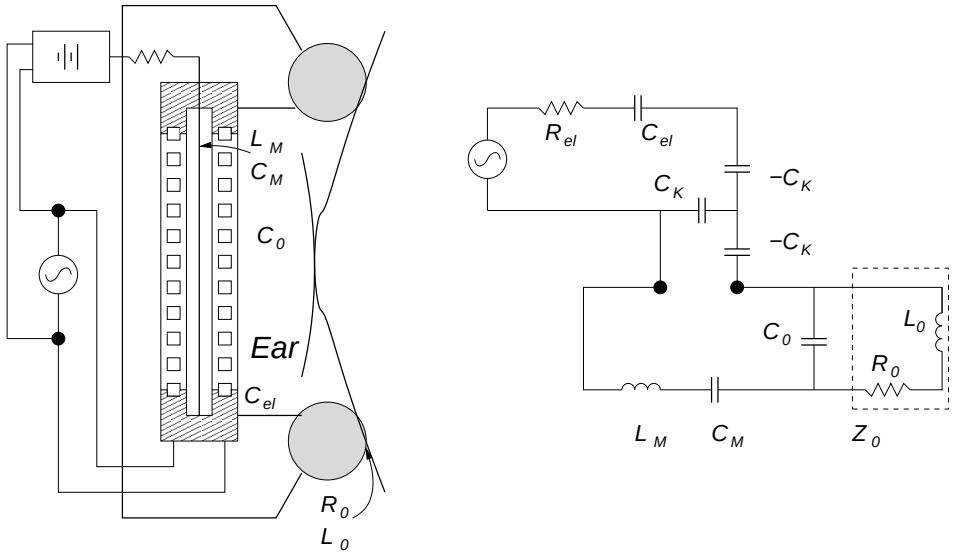

Fig. I.8. Schematized electrostatic headphone (left) and its electroacoustic model (right).

The electroacoustic model, where the acoustic transformer has been substituted by an equivalent sub-network containing two negative capacitances equal to $-C_K$, shows that the equivalent source impedance seen by the acoustic system contains a negative reactive component. In fact, it is easy to calculate that this source impedance is equal to $R_{el} - j\omega C_K^2/C_{el}$, where C_{el} is the capacitance of the electrodes. This reactive component has the desirable property of lowering the frequency peak coming from the usual resonant circuit given by the volume compliance in series with the membrane parameters. Hence, all the considerations made for the isodynamic headphone design apply here, including the flat response (mainly as a result of the lightness of the membrane).

The electret transducer is quite similar to the electrostatic transducer, where the membrane is permanently polarized.

I.3.3 Acoustic load on the ear

The acoustic load of the eardrum on the headphone, and its variations due to pressure changes at the Eustachian tube, are recognized to have effects of low significance on the performance of the headphone system. Conversely, experiments show that the impedance given by ear occlusion caused by the headphones is much more varying in frequency than the equivalent impedance seen by the eardrum in open ear conditions (recall the models shown in Figure I.4). In [163], a certain degree of correlation is shown to exist between the amplitude of these variations and the quality of the equipment, the highest and most unpredictable load being caused by some Walkman headphones under test. The increased OHL experienced using open headphones has been already discussed in section I.2.4.

I.4 Equalization of headphones

In the previous sections we have seen the most important psychoacoustical issues related with headphones, and possible ways to determine and control their low-frequency response according to the type of transducer and ear coupling. It is clear that even a very careful design cannot solve two important questions. First, how to decode the stereophonic material to recreate conditions of binaural listening. Second, how to achieve a high degree of control of the sound pressure at the ear-canal entrance, to reproduce all the cues necessary to give definite source localization, OHL etc.

These two questions are quite closely related. Assuming that the audio material somehow has binaural cues encoded in it, in principle both questions would be solved by decoding and then equalizing the audio signal, both processes now made possible by means of digital signal processing. In practice, and due to the not yet perfectly understood origin of many cues, the use of "transparent" headphones can make equalization much easier and effective. These kind of headphones have been recently made available in the form of open-canal tube-phones, although only for research purposes.

I.4.1 Stereophonic and binaural listening

In the music industry, the recording techniques are not standardized, and many aspects which characterize a recording session are left to the creativity of the professionals working in that field ("Tonmeisters"). Any decoding process devoted to recovering binaural cues from stereophonic material would in principle be ineffective, or even corrupt the quality of sound. Thus, in commercial Hi-Fi equipment stereophonic material feeds the headphones without passing through any special decoding process devoted to extraction of binaural cues. In this case stereophonic listening is performed.

On the opposite side, binaural recording must observe severe requirements. For example, it implies the use of mannequins whose heads host special in-ear microphones. Binaural recording gives quite good results when this material is auditioned using a system for stereophonic listening [113]. Unfortunately, this recording technique has not achieved popularity in the studio, due to the unpredictable artifacts introduced by binaural reproduction when the listening conditions are not precisely determined (see section I.2.2).

Until some years ago, international standards suggested that the frequency response of a headphone had to approximate the flat, free-field response of a loudspeaker located in front of the listener, i.e. the HRTF calculated in correspondence of the frontal position. This choice often resulted in unnatural colouring of sound, without satisfactory frontal localization (see section I.2.2 and I.2.3). This evidence let the manufacturers design a number of alternative frequency responses, from free-field, to diffuse field or even different characteristics [113].

I.4.2 Conditions for binaural listening

As discussed above, binaural cues cannot be easily extracted by stereophonic material.

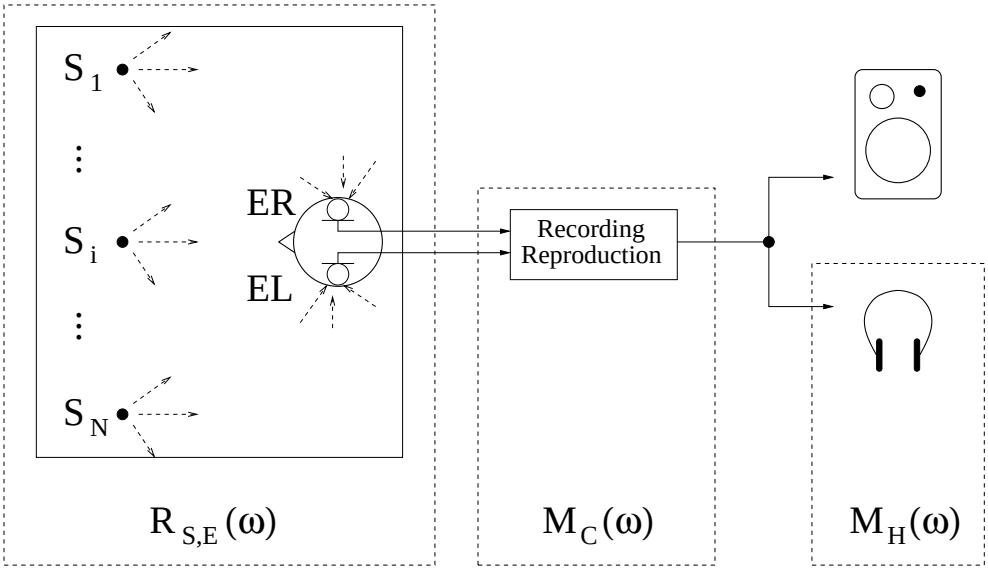

Fig. I.9. Recording and reproduction of binaural material.

N signals $x_{\mathrm{S1}}, \ldots, x_{\mathrm{SN}}$ emitted by sources located at points S1, ..., SN, binaurally recorded using two microphones at positions EL and ER, are then presented to the final reproduction stage as signals $x_{\mathrm{EL}}, x_{\mathrm{ER}}$ given by

$$\begin{vmatrix} X_{\mathrm{EL}}(\omega) \\ X_{\mathrm{ER}}(\omega) \end{vmatrix} = \begin{vmatrix} H_{\mathrm{S1,EL}}(\omega) & \cdots & H_{\mathrm{SN,EL}}(\omega) \\ H_{\mathrm{S1,ER}}(\omega) & \cdots & H_{\mathrm{SN,ER}}(\omega) \end{vmatrix} \begin{vmatrix} X_{\mathrm{S1}}(\omega) \\ \cdots \\ X_{\mathrm{SN}}(\omega) \end{vmatrix} \tag{I.1}$$

or

$$\mathbf{X}_{\mathrm{E}}(\omega) = \mathbf{H}_{\mathrm{S,E}}(\omega)\mathbf{X}_{\mathrm{S}}(\omega) \tag{I.2}$$

where $H_{\mathrm{S}i,\mathrm{E}P}(\omega)$ is a transfer function accounting for all the (desired and undesired, anyway supposed linear) contributions added to the audio signal from its emission by the i-th sound source to the final reproduction, with $P = \{\mathrm{L}, \mathrm{R}\}$.

Binaural information can be extracted from $\boldsymbol{x}_{\mathrm{E}}$ if the components in $\mathbf{H}_{\mathrm{S,E}}$ accounting for all the artifacts introduced by the recording and reproduction chain can be separated, i.e. if a 2×2 matrix $\mathbf{M}_{\mathrm{C}}$ containing all the artifacts exists such that

$$\mathbf{M}_{\mathrm{C}}(\omega)\mathbf{R}_{\mathrm{S,E}}(\omega) = \mathbf{H}_{\mathrm{S,E}}(\omega) \tag{I.3}$$

$\mathbf{R}_{\mathrm{S,E}}$ containing the information about the $2 \times N$ Room Transfer Functions (RTF) calculated from the sound source positions to the recording points plus the binaural cues (see Figure I.9). Headphone listening introduces new artifacts, possibly embedded in another matrix of transfer functions calculated from the headphone input to the ear-canal entrance, written $\mathbf{M}_{\mathrm{H}}$, such that the sound pressure $\boldsymbol{x}_{\mathrm{H}}$ measured at the ear-canal entrances obeys equation

$$\mathbf{X}_{\mathrm{H}}(\omega) = \mathbf{M}_{\mathrm{H}}(\omega)\mathbf{X}_{\mathrm{E}}(\omega) \tag{I.4}$$

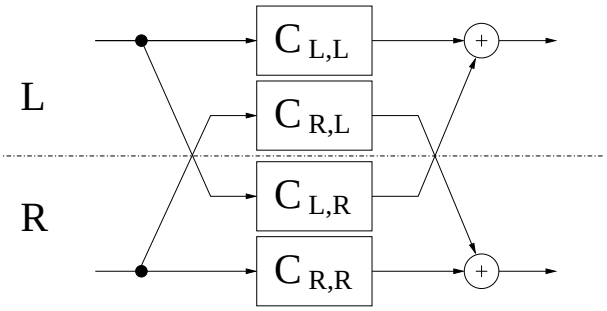

Fig. I.10. Block scheme of a blending circuit.

Binaural listening, in the form of a signal $\boldsymbol{x}_{\mathrm{B}}$ at the ear-canal entrances, would be realized by post-filtering the sound material in order to compensate the artifacts coming from the audio chain, $\mathbf{M}_{\mathrm{C}}$, and the frequency distortions coming from headphones, $\mathbf{M}_{\mathrm{H}}$:

$$\mathbf{X}_{\mathrm{B}}(\omega) = \mathbf{M}_{\mathrm{H}}(\omega)\big\{\mathbf{M}_{\mathrm{C}}(\omega)\mathbf{M}_{\mathrm{H}}(\omega)\big\}^{-1}\mathbf{X}_{\mathrm{E}}(\omega) = e^{-j\omega\tau}\mathbf{R}_{\mathrm{S,E}}(\omega)\mathbf{X}_{\mathrm{S}}(\omega) \tag{I.5}$$

where $e^{-j\omega\tau}$ accounts for an overall delay coming from the compensation of non-minimum phase components [83].

The exact application of these results, i.e. a post-processing system performing a multi-channel equalization equal to $\mathbf{C} = \mathbf{M}_{\mathrm{H}}^{-1}\mathbf{M}_{\mathrm{C}}^{-1}$, is quite never realized because the knowledge of $\mathbf{M}_{\mathrm{H}}$ is often unprecise even in the laboratory, especially at higher frequencies. Nevertheless, approximate solutions of this problem have been attempted, leading to various implementations. In particular, $\mathbf{C}$ can be approximated as a blending stage (see Figure I.10) where the stereophonic signal is mixed to account for the interaural cues occurring during loudspeaker reproduction, otherwise lost if the same material is listened using headphones. Several types and designs of *blending circuits* have been tried, from passive to active networks, even using tubes as acoustic delay lines for the cross-talk signals, but the results obtained using these blending architectures do not justify their systematic introduction in the reproduction chain [113].

I.4.3 Binaural listening with insert earphones

From equation (I.5), one can argue that the easier the compensation is, the higher the expectation to reproduce the binaural signal $\boldsymbol{x}_{\mathrm{B}}$. Recently, technology has made available special earphones [76] which claim to have flat response at the ear-canal entrance, this corresponding to have $|\mathbf{M}_{\mathrm{H}}(\omega)| \approx \boldsymbol{I}$, $\boldsymbol{I}$ being the 2×2 identity matrix. Adopting an audio chain where the response of each stage is linear[3], i.e., assuming $|\mathbf{M}_{\mathrm{C}}(\omega)| \approx \boldsymbol{I}$, one can conclude that reliable conditions for a correct reproduction of binaural recorded material now hold.

If, in particular, the recording session is performed in an anechoic and silent environment using only one sound source located at a specific distance from the

[3] this assumption being quite affordable today

mannequin, then $\mathbf{R}_{\mathrm{S,E}}$ contains the two HRTFs calculated with respect to the specific position of the sound source.

In [87] these peculiar conditions have been recreated to investigate the spectral detail needed for a precise localization of the sound source, using white noise as a test signal, $\boldsymbol{x}_{\mathrm{T}}$. One key of the experiment was that the listener could not recognize whether the sound came from the insert earphones or from one of the loudspeakers inside the chamber, thanks to the insert headphones. This way, the equivalence of loudspeaker and earphone reproduction has been confirmed—white noise in the latter case being properly post-processed in such a way that $\boldsymbol{x}_{\mathrm{B}} = \mathbf{r}_{\mathrm{S,E}} * \boldsymbol{x}_{\mathrm{T}}$. Surprisingly, other valuable conclusions came out from this experiment:

1. reproduction using earphones gave good OHL as well;
2. smoothing the magnitude response in the HRTFs had no consequence until heavy distortions were introduced in the response;
3. corrupting the phase response in the HRTFs (but preserving, of course, mutual ITD), i.e., substituting the HRTFs with their minimum phase versions, did not change the listeners' judgement.

This test would show that OHL is perhaps the most important condition for experiencing consistent binaural cues, and that the sensation of openness (thanks to the negligible acoustic load on the ear and occlusion effect given by insert earphones) is very important to achieve OHL[4]. Once OHL holds, ITD and major spectral details in the HRTFs would give localization cues, unless they are heavily corrupted.

I.5 Conclusions

The psychoacoustics of binaural listening has given a reliable starting point to investigate several questions left unsolved by stereophonic encoding of sound material and headphone listening using common Hi-Fi equipment. Among the most important questions there are the ineffective frontal localization and OHL cues provided by commercial headphones.

Localization and OHL cues can be recovered under certain conditions from binaurally recorded material, using special equipment which bypasses the complicate (and not yet perfectly understood) problem of how to cancel all the artifacts introduced by the recording and reproduction chain. Head-tracking would give help in avoiding further incoherence in the auditory messages, detected when these messages are processed by higher-level cognitive stages [57].

This kind of equipment is available only in the research laboratory at the moment. Despite this, several results in controlled listening conditions would suggest that headphones will be capable of performing binaural reproductions in the near future.

[4] It must be noticed that white noise in general affords a better impression of openness. Moreover, the sensation of openness inside an anechoic chamber is not the same one as in everyday listening conditions.

J

Computation of Linear Filter Networks Containing Delay-Free Loops, with an Application to the Waveguide Mesh

Federico Fontana
IEEE Trans. Speech and Audio Processing.
- ACCEPTED FOR PUBLICATION -

A method that computes linear digital filter networks containing delay-free loops is proposed. Compared with alternative techniques existing in the literature, it does not require a rearrangement of the network structure. Since this method involves the computation of matrices accounting for the connections between the filter blocks forming the network, it becomes computationally interesting when those filters are densely interconnected through delay-free paths. The Triangular Waveguide Mesh is an example of such networks. Using the proposed method, we could compute a transformed version of that mesh containing delay-free loops, obtaining simulations that are significantly more accurate compared with those obtained using the traditional formulation of the triangular mesh.

J.1 Introduction

The *delay-free loop problem* [98, §6.1.3] appears during conversion to the digital domain of some analog filter networks, or during digital-to-digital domain mappings that convert unit delays into blocks that instantaneously respond to the input [33, 81]. As a consequence of that problem, filter networks containing feedback loops may become non-computable after the conversion, since the computation of those loops in the discrete-time cannot be executed any longer by a sequence of operations, due to the lack of pure delays along the loop.

In the case of linear filter networks delay-free loops can be detected and, then, replaced by alternative structures, to obtain equivalent, computable realizations [153]. Such replacements usually result in changes in the original topology of the filter network.

There are cases in which a rearrangement in the filter network topology is deprecated. This happens when a filter structure cannot be merely seen as one particular realization of a transfer function, since it contains essential information about the original system structure. We deal in particular with sound synthesis

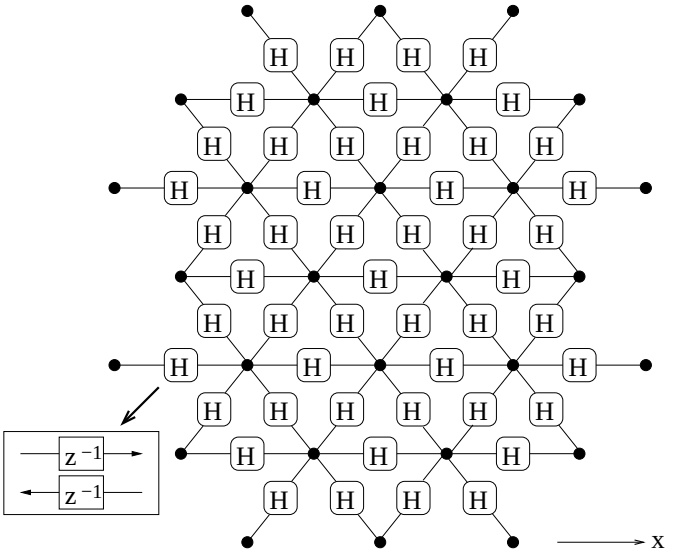

Fig. J.1. Triangular Waveguide Mesh. Adjacent nodes are connected each other via bidirectional unitary delay lines (embedded within the blocks labeled with H, see particular), that receive/inject signals from/to the nodes, respectively. Signals are instantaneously scattered by each node.

using physical models: in that field, it is common to model spatially distributed mechanical and fluid-dynamic systems by means of filter networks that allow to inject energy and acquire responses in particular points of the systems, along with enabling an individual control of the local parameters. The structure of those networks, thus, has a direct correspondence with the distribution in space of a physical system [13, 149].

Recent literature pointed out that a single delay-free loop containing one linear filter in feedback can be computed without rearranging the loop structure [69]. This result has been successfully applied to the computation of Warped IIR filters [70] and magnitude-complementary Tunable Equalization Filters [49].

In this paper, that approach to the delay-free loop computation has been generalized to linear filter networks containing any configuration of delay-free loops. This generalization yields a method that enables to detect the delay-free loops along with computing, at each time step, the signals that are present at any branch of the network.

In the second part of the paper, as an application case we apply the method to the Triangular Waveguide Mesh (TWM, see Figure J.1), a numerical scheme belonging to the family of Digital Waveguide Networks that models the ideal propagation of waves along a two-dimensional homogeneous medium [13, 50, 158]. In that scheme, during each time step, signals incoming to a node are instantaneously scattered out into signals outgoing from the node. As the following time step is triggered, those outgoing signals are forwarded to the adjacent nodes so that they become new incoming signals to be scattered by those nodes.

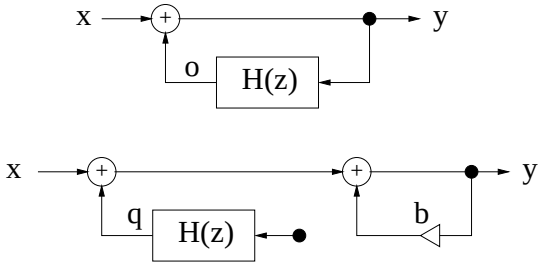

Fig. J.2. Delay-free loop structure (above). Rearrangement of the delay-free loop structure according to the procedure proposed in Section J.2 (below).

The TWM can be seen as a linear filter network in which adjacent scattering nodes are connected one to the other via a bidirectional unitary delay line (embedded, in Figure J.1, within blocks labeled with H). With that scheme, a spectral transformation F mapping each unit delay into a corresponding first-order allpass transfer function A, i.e., the following transformation of the z-variable [98]:

$$z^{-1} = F^{-1}(\tilde{z}) = A(\tilde{z}) = \frac{\tilde{z}^{-1} - \lambda}{1 - \lambda\tilde{z}^{-1}} \quad , \tag{J.1}$$

has been shown to have considerable benefits in the accuracy of the results provided by the model, once the allpass coefficient λ is properly set [137].

The proposed method allows to compute the signal in the TWM after the network has been transformed using (J.1). Responses coming from the "warped" version of the TWM, from here called warped TWM, are presented in the conclusion of the paper.

J.2 Computation of the Delay-Free Loop

First, we briefly revise the method proposed by Härmä [69] for computing a single delay-free loop in the form expressed by the upper structure in Figure J.2, where x and y are respectively the input and output signals, and it is

$$H(z) = \frac{b + \sum_{k=1}^{M} b_k z^{-k}}{1 - \sum_{k=1}^{N} a_k z^{-k}} \quad . \tag{J.2}$$

The method considers, at any discrete-time step n, the output o from the feedback branch as a linear superposition of two components: the former accounts for the instantaneous reaction of the filter in feedback; the latter, q, depends on its internal state. Thus, by (J.2), it is $o[n] = by[n] + q[n]$. Hence, the output from the delay-free structure can be written as

$$y[n] = x[n] + o[n] = x[n] + by[n] + q[n] \quad , \tag{J.3}$$

which implies

$$y[n] = \frac{1}{1-b}\left\{x[n] + q[n]\right\} \quad . \tag{J.4}$$

This algebraical rearrangement of the delay-free loop can be always done, even for unstable structures, except for cases in which it is $b = 1$. In those cases it can be demonstrated that the structure is non-causal, since its input/output transfer function has a singularity in the infinite point of the z-plane [69].

Since q can be computed at any time step by feeding the filter with a null sample, then the output from the structure can be computed by means of a procedure that makes use of three steps, which are executed at each n (see the lower part of Figure J.2):

1. q is computed feeding the feedback filter with zero;
2. y is computed using (J.4);
3. y is fed to the filter in feedback, to update its state correctly.

J.3 Computation of Delay-Free Loop Networks

Any procedure leading to the computation of the delay-free loop must, in a way or another, rearrange the non-computable structure. The procedure proposed in Section J.2 isolates a sub-loop that accounts for the instantaneous part of the response of the feedback filter (see Figure J.2). That idea is now generalized to linear filter networks containing multiple delay-free loops.

Suppose to isolate, from the filter network, a number of subnetworks containing delay-free loops. In order to do this we can take advantage from a graph-theoretic detection method existing in the literature [153]. We now treat each subnetwork separately.

Let us index each one of the L filter branches forming a subnetwork. Each branch provides the transfer function

$$H_i(z) = \frac{\displaystyle\sum_{k=0}^{M_i} {b_k}^{(i)} z^{-k}}{\displaystyle 1 - \sum_{k=1}^{N_i} {a_k}^{(i)} z^{-k}} \quad , \quad i = 1, \ldots, L \quad . \tag{J.5}$$

Similarly to what we have seen in Section J.2, for each one of those branches we isolate, in the output y_i, the addendum containing the instantaneous input x_i. Hence, for the i-th filter branch we can write

$$y_i[n] = b_i x_i[n] + q_i[n] \quad , \tag{J.6}$$

where clearly it is $b_i = {b_0}^{(i)}$. Note that b_i is non null, since the subnetwork resulting after the application of the delay-free loop detection method cannot contain branches corresponding to pure delays. On the other hand, q_i can be null at any time step when the filter in the i-th branch reduces to a multiplier. In particular,

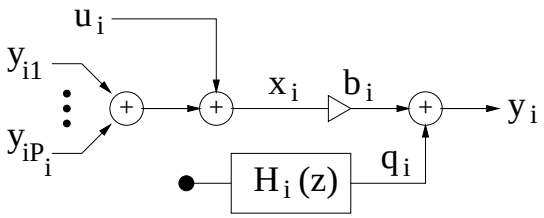

Fig. J.3. Structure of a filter branch. The output y_i is obtained as a superposition of one component depending on the input x_i, plus one input-independent component q_i. The input is the result of summing outputs $y_{i_1}, \ldots y_{i_{P_i}}$ from other branches plus one external input u_i.

we can account for direct connections, i.e., $y_i = x_i$: they will be represented by branches where it is $b_i = 1$ and $q_i \equiv 0$.

Now, we consider the inputs to any filter branch. Each input x_i is the superposition of P_i outputs $y_{i_1}, \ldots y_{i_{P_i}}$ from other branches plus an external input u_i, null if the filter branch is fed only by outputs from other branches (see Figure J.3):

$$x_i[n] = \sum_{k=1}^{P_i} y_{i_k}[n] + u_i[n] \quad . \tag{J.7}$$

Note that direct loopbacks connecting y_i and x_i are avoided. In other words, condition

$$i_k \neq i \ , \ k = 1, \ldots, P_i \tag{J.8}$$

must be satisfied for any i. For this purpose, each direct loopback is indexed in the subnetwork separately, in a way that its index respects condition (J.8).

By collecting the outputs $y_1, \ldots, y_L$ together in the column vector $\boldsymbol{y}$, and doing the same also for x, q and u, (J.6) and (J.7) can be expressed by the following vectorial equations, respectively:

$$\boldsymbol{y}[n] = \boldsymbol{B}\boldsymbol{x}[n] + \boldsymbol{q}[n] \tag{J.9a}$$

$$\boldsymbol{x}[n] = \boldsymbol{P}\boldsymbol{y}[n] + \boldsymbol{u}[n] \tag{J.9b}$$

where $\boldsymbol{B}$ is a non-singular matrix containing the coefficients $b_1, \ldots, b_L$ in its diagonal and zero elsewhere, and $\boldsymbol{P}$ is a matrix accounting for the connections among branches. Element p_{ij} of $\boldsymbol{P}$ is equal to one if the output of the j-th filter branch feeds the i-th filter, otherwise it is equal to zero. The exclusion of direct loopbacks in the network, expressed by (J.8), constrains the diagonal of this matrix to be null.

Substituting (J.9b) in (J.9a), after the elimination of $\boldsymbol{x}$, we find a way to calculate the filter outputs from values that are independent of the delay-free loops:

$$\boldsymbol{F_y}\boldsymbol{y}[n] = \boldsymbol{B}\boldsymbol{u}[n] + \boldsymbol{q}[n] \quad , \tag{J.10}$$

with

$$\boldsymbol{F_y} = \boldsymbol{I} - \boldsymbol{BP} = \begin{pmatrix} 1 & -b_1 p_{12} \ldots & & -b_1 p_{1L} \\ & \ddots & & \\ \vdots & & 1 & \vdots \\ & & & \ddots \\ -b_L p_{L1} & & \ldots -b_L p_{L\,L-1} & 1 \end{pmatrix} \quad , \tag{J.11}$$

$\boldsymbol{I}$ being the $L \times L$ identity matrix. Each element ij of $\boldsymbol{F_y}$ contains the i-th diagonal element of $\boldsymbol{B}$ changed in sign, gated by the element ij of $\boldsymbol{P}$.

The existence of the inverse matrix ${\boldsymbol{F_y}}^{-1}$ determines whether the output $\boldsymbol{y}$ can be calculated or not. Here we show that if the inverse matrix does not exist, then the subnetwork upon consideration contains non-causal loops. In fact, after formal substitution of $\boldsymbol{y}$, $\boldsymbol{u}$ and $\boldsymbol{q}$ with their corresponding z-transforms, respectively $\boldsymbol{Y}$, $\boldsymbol{U}$ and $\boldsymbol{Q}$, then (J.10) can be expressed in the z-domain:

$$\boldsymbol{F_y Y}(z) = \boldsymbol{BU}(z) + \boldsymbol{Q}(z) \quad , \tag{J.12}$$

such that

$$\boldsymbol{Y}(z) = \frac{\mathrm{adj}\boldsymbol{F_y}}{\det \boldsymbol{F_y}} \left\{ \boldsymbol{BU}(z) + \boldsymbol{Q}(z) \right\} \tag{J.13}$$

where the operator adj gives the adjoint matrix.

In the hypothesis that the input signal is causal and bounded, and the initial energy in the network is finite (that is, $|\boldsymbol{q}[0]| < \boldsymbol{\infty}$), then, recalling a known theoretical result [110, §4.4.8], we have

$$\lim_{z \to \infty} \boldsymbol{Y}(z) = \frac{\mathrm{adj}\boldsymbol{F_y}}{\det \boldsymbol{F_y}} \left\{ \boldsymbol{Bu}[0] + \boldsymbol{q}[0] \right\} \quad . \tag{J.14}$$

This limit is equal to infinity for some scalar functions $Y_i(z), i = 1, \ldots, L$, as long as $\det \boldsymbol{F_y}$ goes to zero and it is $\boldsymbol{Bu}[0] + \boldsymbol{q}[0] \neq \boldsymbol{0}$. Hence, if $\boldsymbol{F_y}$ is singular then the subnetwork contains non-causal loops.

In the case when it is $\boldsymbol{u} \equiv \boldsymbol{0}$ and $\boldsymbol{q} \equiv \boldsymbol{0}$ during the whole process, the limit (J.14) cannot be determined when $\boldsymbol{F_y}$ is singular. This case includes pathological situations. For instance, consider the delay-free loop of Figure J.2, where the loopback filter is a pure multiplier (that is, $q \equiv 0$), in a way that the upper and lower structures match each other. Let us set $b = 1$ and $u = 0$ at time step n. In principle, the loop will keep on producing an output that is constantly equal to $y[n-1]$ as long as u is null. Otherwise, i.e., for $b = 1$ and $u \neq 0$, the output cannot be calculated.

On the other hand, stability is not necessarily required for the existence of ${\boldsymbol{F_y}}^{-1}$. The inverse matrix allows to calculate the output explicitly:

$$\boldsymbol{y}[n] = {\boldsymbol{F_y}}^{-1} \boldsymbol{Bu}[n] + {\boldsymbol{F_y}}^{-1} \boldsymbol{q}[n] \quad . \tag{J.15}$$

In summary, at each time step the procedure that computes the subnetwork containing delay-free loops is the following:

1. compute $\boldsymbol{q}[n]$, feeding each filter branch with zero;
2. compute $\boldsymbol{y}[n]$ from (J.15);
3. compute $\boldsymbol{x}[n]$ from (J.9b);
4. feed the filters in the subnetwork with $\boldsymbol{x}[n]$ to update their state.

Symmetrically, another procedure can be executed. In fact, Equation (J.9a) can be substituted in (J.9b) in such a way that, after the elimination of $\boldsymbol{y}$, an explicit solution is found out for $\boldsymbol{x}[n]$ rather than $\boldsymbol{y}[n]$:

$$\boldsymbol{x}[n] = \boldsymbol{F_x}^{-1}\boldsymbol{P}\boldsymbol{q}[n] + \boldsymbol{F_x}^{-1}\boldsymbol{u}[n] \quad , \tag{J.16}$$

with

$$\boldsymbol{F_x} = \boldsymbol{I} - \boldsymbol{P}\boldsymbol{B} \quad . \tag{J.17}$$

Equation (J.16) can be solved if and only if (J.15) admits solution (see Appendix).

Hence, a symmetric procedure can be devised:

1. compute $\boldsymbol{q}[n]$, feeding each filter branch with zero;
2. compute $\boldsymbol{x}[n]$ from (J.16);
3. compute $\boldsymbol{y}[n]$ from (J.9a);
4. feed the filters in the subnetwork with $\boldsymbol{x}[n]$ to update their state.

Finally, it is worth mentioning that the method works also in the case when $\boldsymbol{B}$ is singular, that is, when some filter branches in the subnetwork do not respond instantaneously to the input. In other words, filter branches characterized by having $b_i = 0$ can be included in the subnetwork without harming the procedure. This is proved in the Appendix.

The possibility to extend the given method to any linear filter network is useful, for example, if the filter coefficients are varied runtime, since it prevents the procedure to fail as long as b_i is set to zero at some branches during the computation.

J.3.1 Detection of Delay-Free Loops

The extension to cases in which it is $b_i = 0$ for some branches enables the method to detect the delay-free loops that are present in a filter network. To show this, first we give a rule for deciding which filter branches are part of a delay-free loop. The decision is taken checking the two following conditions for any pair (i, j) of filter branches. Thus, two branches i and j are part of the same delay-free loop if

1. a variation of q_j generates a perturbation in y_i instantaneously, and
2. a variation of q_i generates a perturbation in y_j instantaneously,

otherwise those branches are not part of the same delay-free loop.

Note that the rule holds also in the case when $q_i \equiv 0$ and/or $q_j \equiv 0$, i.e., if branches corresponding to pure multipliers are involved in the decision. In that case, an equivalent variation occurring at the same point where q_i and/or q_j is injected into the loop (refer to Figure J.3) can be considered instead.

The two statements can be checked automatically once $\boldsymbol{F_y}^{-1}$ is known. In fact, it can be seen from (J.15) that this matrix puts the filter outputs in relation

with the state vector, in such a way that the verification of the two statements corresponds to checking that

$$(\boldsymbol{F_y}^{-1})_{ij} \neq 0 \quad , \quad (\boldsymbol{F_y}^{-1})_{ji} \neq 0 \tag{J.18}$$

where $(\cdot)_{ij}$ is the element ij of a matrix. In other words, two branches i and j are part of the same delay-free loop if and only if the element ij of $\boldsymbol{F_y}^{-1}$ and its transposed are not null.

In the Appendix, a delay-free loop network proposed by Szczupak and Mitra [153, example 2] is resolved using the detection method proposed here.

J.4 Scope and Complexity of the Proposed Method

The proposed method involves the definition of a linear system from each subnetwork containing delay-free loops, whose dimension equals the number of filter branches existing in the subnetwork. The solution of that system provides formulas, (J.15) or (J.16), that yield the output $\boldsymbol{y}$ or input $\boldsymbol{x}$, respectively. We cannot avoid the matrix calculations contained in (J.15) or (J.16), since any state variation and/or any new external input occurring somewhere in the subnetwork produce global effects on it in the form of output variations at all the filter branches composing the subnetwork, except for those that do not respond instantaneously to an input (for which the method is useless).

Although digital signal processors are usually well-suited to perform sequential matrix computation, and given that the computation of (J.15) or (J.16) can be parallelized, for example over L processors, yet the key feature of the method is the definition of a procedure that allows to calculate not only the output from the subnetwork, but also its internal signals.

Nevertheless, its computational performance can be interesting in the case of "dense" networks, that is, networks where each filter block feeds most of the other filters. In the limit case when each branch feeds all the other branches, and assuming that all filters in the network are N-th order filters, each one provided with an external input, then we have from Eqs. (J.7) and (J.5):

$$Y_i(z) = H_i(z)X_i(z) = H_i(z)\left\{\sum_{\substack{k=1 \\ k\neq i}}^{L} Y_k(z) + U_i(z)\right\}, \quad i = 1, \ldots, L \quad . \tag{J.19}$$

Equation (J.19) can be expressed in matrix form,

$$\boldsymbol{H}_{-}(z)\boldsymbol{Y}(z) = \boldsymbol{U}(z) \quad , \tag{J.20}$$

with

$$\boldsymbol{H}_{-}(z) = \begin{pmatrix} \frac{1}{H_1} & -1 & -1 & \ldots & -1 \\ -1 & \frac{1}{H_2} & -1 & \ldots & -1 \\ \vdots & & \ddots & & \vdots \\ -1 & \ldots & -1 & \frac{1}{H_{L-1}} & -1 \\ -1 & \ldots & -1 & -1 & \frac{1}{H_L} \end{pmatrix} \quad . \tag{J.21}$$

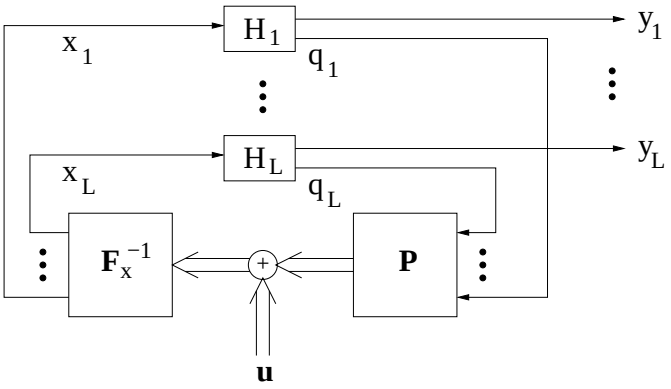

Fig. J.4. Schematic of the latter procedure given in Section J.3. Individual filter blocks are in transposed direct form.

Assuming that the inverse of $\boldsymbol{H}_-$ exists, then we have

$$\boldsymbol{Y}(z) = \boldsymbol{H}_-(z)^{-1}\boldsymbol{U}(z) = \frac{\mathrm{adj}\boldsymbol{H}_-(z)}{\det \boldsymbol{H}_-(z)}\boldsymbol{U}(z) \quad . \tag{J.22}$$

Equation (J.22) shows that in the limit case hypothesized here the extraction of all the filter outputs, using a traditional method that rearranges delay-free loops into an alternative network, involves in principle the computation of L^2 filters having order LN, each one represented by one element of $\boldsymbol{H}_-(z)^{-1}$. In other words, the computation of a dense filter network containing delay-free loops turns into an algorithm that, without a peculiar reformulation of the structure that computes $\boldsymbol{H}_-(z)^{-1}$, in general has a complexity $O(NL^3)$.

Compared to that, the proposed method needs:

- $4(L-1)^2$ operations to compute (J.15) or (J.16)—step 2 of both procedures given in Section J.3;
- $2(L-1)^2 + L$ operations to compute $\boldsymbol{x}$ from (J.9b) or $\boldsymbol{y}$ from (J.9a)—step 3 of the same procedures;
- another order of LN operations to compute $\boldsymbol{q}$ and to update the state of the L filters—steps 1 and 4.

Then, the algorithm coming from the proposed method has a complexity $O(NL^2)$.

In the case when the filters $H_1, \ldots, H_L$ can be realized in transposed direct form, then $\boldsymbol{q}$ can be directly read from the filter states so that step 1 is skipped. A schematic of the latter procedure given in Section J.3 is shown in Figure J.4, where the individual filter blocks have been realized in transposed direct form so that they yield $\boldsymbol{q}$ directly.

In most of the potential application cases $\boldsymbol{F_y}$ and $\boldsymbol{F_x}$ are sparse matrices, since the filter outputs feed only a small part of the inputs belonging to other filter branches. In those cases, efficient inversion procedures can be invoked [65] if $\boldsymbol{F_y}^{-1}$ or $\boldsymbol{F_x}^{-1}$ must be recomputed runtime, for example when the filter coefficients are varied. Although traditional delay-free loop computation methods seem to cope more effectively with those cases, especially in real-time applications, it should be

noted that rearrangements in the filter network, that are needed by those methods to remove the delay-free loops, usually lead to structures where the filter coefficients are no longer directly tunable.

J.5 Application to the TWM

TWMs define stable, energy-preserving filter networks, whose structures have a direct correspondence with the geometry of an ideal elastic membrane [13]. The scattering nodes have also the function of input/output points, so that they form a discrete set of locations where the (position-dependent) transfer functions measured from one input point to one output point of the membrane can be computed from the model (see Figure J.1).

Waveguide Mesh models introduce, in the simulations, a numerical error called *dispersion* [152, 158]. In the case when a TWM is used to model a membrane, then dispersion can be equivalently interpreted, with good approximation, as a frequency-dependent change in propagation speed of the waves that resonate in the system, causing a proportional misplacement of the modal frequencies in the model response. This change in speed is quantified by the dispersion function D, that yields the propagation speed ratio between modeled ("dispersed") and ideal ("undispersed") traveling wave components. The dispersion function decreases with frequency, suggesting that high frequency waves are affected by a larger propagation speed error [50].

A manipulation of the wave propagation speed would have benefits in the accuracy of the simulations provided by the TWM. This manipulation can be done by transforming each pure delay into a frequency-dependent phase shift, in a way that the new blocks obtained after that transformation reduce the error caused by dispersion. In practice, these blocks introduce a specific phase delay to any frequency component as long as it is transferred from one scattering node to another.

Allpass-to-allpass transformations may provide that kind of manipulation if their parameters are properly set, meanwhile they preserve the properties of stability and energy preservation of the TWM. The spectral transformation F given by (J.1) has been shown to improve the model accuracy significantly [137]. Previous literature proposed *frequency warping* as a technique for computing off-line the TWM as if it were transformed by F, along with a strategy for optimizing the λ coefficient over a limited part of the spectrum [137]. We exploit the method introduced in Section J.3 to compute the transformed TWM online, optimizing λ over the entire spectrum.

The following instantaneous scattering equation holds for any internal node of a TWM [150]:

$$\begin{vmatrix} x_{A-} \\ x_{B-} \\ x_{C-} \\ x_{D-} \\ x_{E-} \\ x_{F-} \end{vmatrix} = \begin{pmatrix} \frac{-2}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{-2}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{-2}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{-2}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{-2}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{-2}{3} \end{pmatrix} \begin{vmatrix} y_{A+} \\ y_{B+} \\ y_{C+} \\ y_{D+} \\ y_{E+} \\ y_{F+} \end{vmatrix} \tag{J.23}$$

where $A+, \ldots, F+$ and $A-, \ldots, F-$ index the branches incoming to and outgoing from the internal scattering node, respectively.

The nodes forming the boundary reverse the sign of any incoming wave signal, given that the TWM simulates a membrane ideally clamped at the border:

$$x_{G-} = -y_{G+} \quad , \tag{J.24}$$

where $G+$ and $G-$ are indexes of filter branches incoming to and outgoing from the boundary node, respectively.

Relations (J.23) and (J.24) are used to form the matrix $\boldsymbol{P}$, that, hence, has the aspect of a sparse matrix: elements accounting for boundary reflections are equal to -1; elements accounting for internal wave reflections are equal to $-2/3$; finally, elements accounting for internal wave transmissions from one branch to another are equal to $1/3$. Note that a rigorous application of the method, i.e., the definition of a matrix $\boldsymbol{P}$ containing only elements equal to either zero or one, would have required to define a larger number of filter branches with no advantages in the implementation of the model.

The output signal v from a node can be computed, at each time step, doubling the average of either its incoming or outgoing wave signals:

$$v = \frac{1}{3}\{y_{A+} + \ldots + y_{F+}\} = \frac{1}{3}\{x_{A-} + \ldots + x_{F-}\}\,. \tag{J.25}$$

from (J.1) it is $\boldsymbol{B} = -\lambda \boldsymbol{I}$. From (J.11) and (J.17) we have $\boldsymbol{F_y} = \boldsymbol{F_x} = \boldsymbol{I} + \lambda \boldsymbol{P}$. $\boldsymbol{F_y}^{-1}$ and $\boldsymbol{F_x}^{-1}$ do not contain null elements, suggesting that the warped TWM is made of one delay-free loop as a whole, where a variation occurring at any filter branch induces perturbations over the whole rest of the mesh. This behavior must not lead the reader to the conclusion that wave signals in the warped TWM propagate at *infinite* speed along the mesh. Rather, it can be seen as an effect of the finite bandwidth of the discretized wavefronts propagating along the warped TWM.

If the allpass filters reduce to pure delays ($\lambda = 0$), nevertheless the computing procedure keeps on working as explained at the end of Section J.3. Under this condition it is $\boldsymbol{B} = \boldsymbol{0}$, and $\boldsymbol{F_y} = \boldsymbol{F_x} = \boldsymbol{I}$. In that case it can be easily seen that, during the free evolution of the system, both procedures given in Section J.3 reproduce exactly the computations that are required to simulate the free evolution of the original TWM [50, 158].

J.5.1 Reducing dispersion

Dispersion can be exactly calculated for any (two-dimensional) spatial frequency component traveling along the mesh [13, 152]. In polar coordinates, dispersion can be seen as a function of direction θ and magnitude ξ of the spatial frequencies, $D(\theta, \xi)$.

In the TWM, traveling wave components propagate with a speed error that is, with good approximation, independent of the direction of propagation [137]. For this reason, dispersion can be averaged into a single-variable function $D(\xi)$ that, although depending only on the frequency magnitude, still provides information with sufficient precision:

$$D(\xi) = \frac{1}{2\pi}\int_0^{2\pi} D(\theta,\xi)d\theta = \frac{2}{\pi}\int_0^{\pi/2} D(\theta,\xi)d\theta \quad . \tag{J.26}$$

This averaging can be safely done up to the edge of the spatial bandwidth, that is equal to $\xi_{\mathrm{MAX}} = 1/(d\sqrt{3})$ in the case of a TWM whose scattering nodes are separated by a distance d, corresponding to the waveguide length [53].

Recalling the nominal (undispersed) wave propagation speed in a two-dimensional waveguide mesh [158]

$$c = \frac{1}{\sqrt{2}}dF_s \quad , \tag{J.27}$$

where F_s is the sampling frequency, then from $D(\xi)$ we can calculate the misplacement affecting a resonance frequency f. In fact, that resonance comes from a standing wave having spatial frequency magnitude $\xi = f/c$, in a way that the dispersion ratio for that resonance is equal to:

$$\frac{\overline{f}}{f} = D(\frac{f}{c}) = D(\frac{f\sqrt{2}}{dF_s}) \quad , \tag{J.28}$$

where $\overline{f}$ is the frequency of the misplaced resonance. This formula is valid over a frequency domain limited by

$$f_{\mathrm{MAX}} = c\xi_{\mathrm{MAX}} = \frac{1}{\sqrt{2}}dF_s\frac{1}{d\sqrt{3}} = \frac{F_s}{\sqrt{2}\sqrt{3}} < \frac{F_s}{2} \quad . \tag{J.29}$$

Above this limit (that lies below the Nyquist limit, as emphasized also in (J.29)) the TWM does not generate resonances[1].

From (J.28) we can immediately figure out a function $\mathcal{D}$, mapping undispersed into dispersed frequencies:

$$\overline{f} = \mathcal{D}(f) = fD(\frac{f}{c}) \quad . \tag{J.30}$$

On the other hand, the transformation (J.1) induces the following mapping between the domain of the untransformed (i.e., misplaced) resonances $\overline{f}$ and the transformed (or *warped*) frequencies $\tilde{f}$ [101]:

$$\tilde{f} = \mathcal{F}(\overline{f}) = \frac{F_s}{2\pi}\arctan\frac{(1-\lambda^2)\sin(2\pi\overline{f}/F_s)}{2\lambda + (1+\lambda^2)\cos(2\pi\overline{f}/F_s)} \tag{J.31}$$

where it is $\tilde{z} = e^{i2\pi\tilde{f}/F_s}$. In particular, the derivative calculated at dc of (J.31) is equal to

$$\mathcal{F}_0 = \left\{\mathcal{F}(\overline{f})\right\}'_{\overline{f}=0} = \frac{1-\lambda}{1+\lambda} \quad . \tag{J.32}$$

[1] More precisely, the TWM can process two-dimensional signals having spectral components up to $\xi = 2/(3d)$, that is, recalling (J.29), the temporal frequency domain is *absolutely* limited by $2c/(3d) = F_s\sqrt{2}/3 \approx 0.471F_s$. Such components exist only for certain directions θ [53]: for this reason, the spatial band ranging from ξ_{MAX} to $2/(3d)$ cannot be part of the domain of $D(\xi)$, although resonances located at frequencies that are higher than f_{MAX} appear in the impulse response of a TWM.

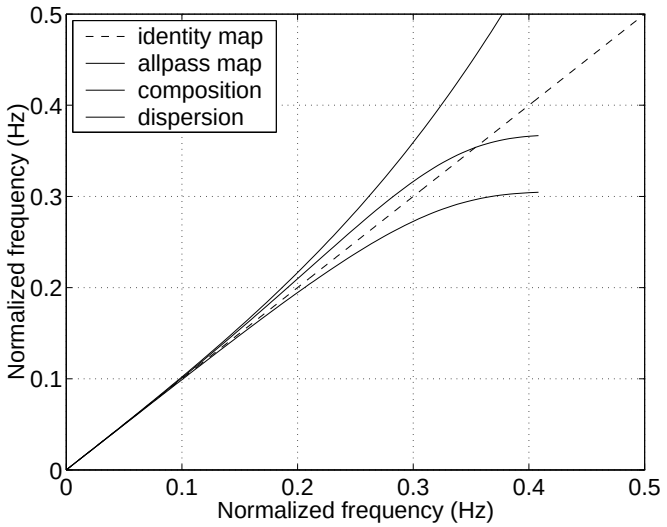

Fig. J.5. Normalized frequency domain mappings induced by the warped TWM (solid lines) plus identity map to be taken as reference (dashed line). Referring to the right end of each plot: dispersion function $\mathcal{D}$ (bottom); allpass map $\mathcal{F}/\mathcal{F}_0$ with $\lambda = 0.2200$ (top); composition of the two (middle).

Equation (J.32) gives reason of the fact that dispersion is reduced by the allpass map (J.1) at the cost of an overall phase delay change (also called *warping ratio* in the literature) that, for positive values of λ, shrinks the frequency response by a factor $\mathcal{F}_0$ [137]. This corresponds to reducing the wave propagation speed in the TWM by the same factor.

Summarizing, in the warped TWM the frequency domain is transformed by two maps, the former caused by dispersion, the latter induced by the allpass transformation:

$$f \quad \stackrel{\mathcal{D}}{\longrightarrow} \quad \overline{f} \quad \stackrel{\mathcal{F}}{\longrightarrow} \quad \tilde{f} \quad , \tag{J.33}$$

such that

$$\tilde{f} = \mathcal{F}\{\mathcal{D}(f)\} \qquad , \qquad 0 \leq f < f_{\mathrm{MAX}} \tag{J.34}$$

An optimal value for λ can be found, for instance, minimizing the distance between $\tilde{f}$ and f over the domain where the TWM produces resonances, independently of the warping ratio $\mathcal{F}_0$:

$$\min_{\lambda:|\lambda|<1} \left|\left| \frac{\tilde{f}}{\mathcal{F}_0} - f \right|\right|_2 \qquad , \qquad 0 \leq f < f_{\mathrm{MAX}} \quad . \tag{J.35}$$

We have computed (J.26), (J.34) and (J.35) numerically, assuming as a metric for distance the L_2 norm of a real sequence s, i.e., $||s||_2 = \sum_l s^2[l]$. The above minimization yields $\lambda = 0.2200$, and, consequently, $\mathcal{F}_0 = 0.6393$. Figure J.5 shows (in normalized frequency scales) plots of both maps, $\mathcal{D}$ and $\mathcal{F}$ (the latter divided by $\mathcal{F}_0$ to hide domain shrinking), together with their composition using the optimal value λ.

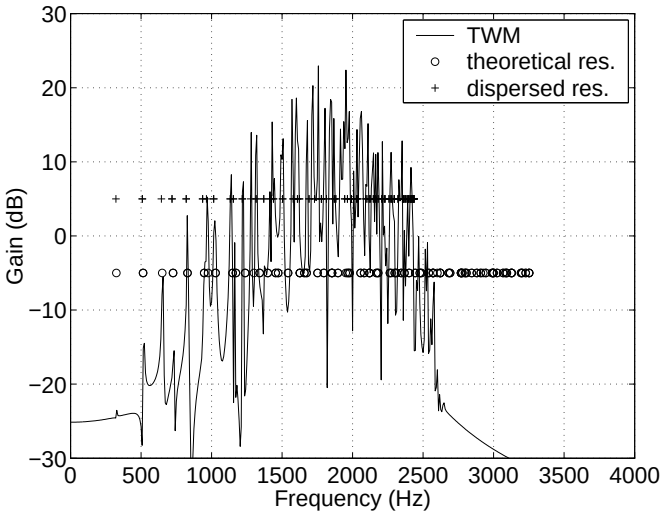

Fig. J.6. Frequency response ($F_s = 8$ kHz) of a TWM modeling a square membrane sized 0.3 m; waves traveling at $c = 131$ m/s. Resonance positions of the membrane below f_{MAX} ('$\circ$'). Dispersed resonance positions of the membrane below f_{MAX} ('+').

J.5.2 Results

Figure J.6 shows the frequency response coming from a TWM model, sampled at 8 kHz, of a square membrane sized 0.3 m, with waves traveling at a speed equal to 131 m/s. Those quantities define a TWM having size equal to 13 waveguides (numbering them along one of the waveguide orientations, such as x in Figure J.1). Together with the plot of the response, the positions of the theoretical modes resonating in that membrane below f_{MAX} are depicted with '$\circ$' (without dispersion) and '+' (after dispersion), respectively. The correspondence between theory and simulation is good.

Some interesting considerations emerge from the analysis of that response. Since the excitation and the acquisition points are located close to the boundary, the lowest resonances (in particular the first one, called the *fundamental*) have a smaller magnitude. Moreover, the magnitude response drops down above $\overline{f}_{\mathrm{MAX}}$, i.e., above the frequency labeled by the upmost '+' symbol[2].

This TWM is made of 998 waveguide branches. Thus, the computation of the corresponding warped TWM involves matrices having identical dimension. Figure J.7 shows figures taken from that warped TWM, assuming $\lambda = 0.2200$. The matching between theoretical resonance positions (depicted with '+'), calculated for the warped model using (J.34), and corresponding peak positions in the warped TWM response is satisfactory. Moreover, the theoretical modes produced by the original square membrane below f_{MAX} have been presented once more, using '$\circ$' symbols, after their positions have been divided by $\mathcal{F}_0$ to account for the proportional wave speed reduction caused by the allpass transformation.

[2] As briefly explained in footnote 1, this happens because only a few, direction-dependent waves can resonate in the TWM beyond that frequency.

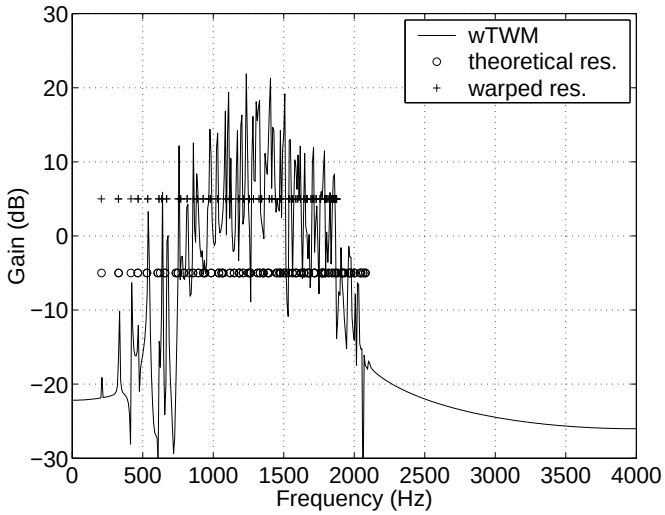

Fig. J.7. Frequency response ($F_s = 8$ kHz) of a warped TWM ($\lambda = 0.2200$) modeling a square membrane sized 0.3 m; waves traveling at $c = 131$ m/s. Resonance positions of the membrane below f_{MAX} corrected by factor $\mathcal{F}_0 = 0.6393$ ('○'). Calculated resonance positions for the warped TWM ('+').

Alternative minimizations can be chosen instead of (J.35), for instance operating on certain bands such as the low frequency [137]. Moreover, higher order allpass-to-allpass transformations can be taken into consideration: although resulting in more accurate phase equalizations, meanwhile they might excessively distort the overall frequency response. As a starting point, the warping ratio can be used as a prior figure for the analysis of the overall distortion caused by a transformation.

The proposed method can be used in alternative configurations of Digital Waveguide Networks, whenever an improvement in the accuracy of the simulations justifies its application [13, 51, 138].

J.6 Conclusion

A method for the detection and computation of delay-free loops in linear filter networks has been proposed. That method translates automatically into a procedure that detects the delay-free loops and computes such networks. Compared with traditional methods, the efficiency of the computation becomes interesting in the case of dense filter networks, as long as their branches process signals that must be presented at the system output.

The proposed method turns out to be useful in the simulation of distributed systems, such as Digital Waveguide Networks, when they implement delay-free transmission in consequence of the application of spectral mappings that eliminate pure delays. As an application case we have computed a warped Triangular

Waveguide Mesh, that has been shown to provide more accurate simulations compared with the traditional triangular mesh.

Acknowledgments

The Author is grateful to Dr. Aki Härmä for his insightful and patient feedback, to Prof. Davide Rocchesso, who helped with several suggestions, and to Prof. Matti Karjalainen, who contributed to inspire this research.

J.7 Appendix

J.7.1 Existence of the solutions

Recalling that $\boldsymbol{B}$ is full-rank, it is

$$\begin{aligned} \det \boldsymbol{F_x} &= \det(\boldsymbol{I} - \boldsymbol{PB}) = \det(\boldsymbol{B}^{-1} - \boldsymbol{P}) \det(\boldsymbol{B}) \\ &= \det(\boldsymbol{B}) \det(\boldsymbol{B}^{-1} - \boldsymbol{P}) = \det(\boldsymbol{I} - \boldsymbol{BP}) = \det \boldsymbol{F_y} \quad . \end{aligned} \tag{J.36}$$

J.7.2 Extension of the method to any linear filter network

Let $b_i = 0$ for the i-th filter branch. Recalling (J.11), this means that the i-th row of $\boldsymbol{F_y}$ is null except for the i-th element, that is equal to one. Hence, we have that

$$\det \boldsymbol{F_y} = (\det \boldsymbol{F_y})_i \quad , \tag{J.37}$$

where $(\det \boldsymbol{F_y})_i$ is the minor determinant resulting from the removal of the i-th row and column from $\det \boldsymbol{F_y}$. By (J.36), this means that Eqs. (J.15) and (J.16) admit solution in particular when it is $b_i = 0$.

Relation (J.37) can be repeatedly applied to all the branches, say $i_1, \ldots, i_R$, such that $b_{i_1} = \ldots = b_{i_R} = 0$:

$$\det \boldsymbol{F_y} = (\ldots(((\det \boldsymbol{F_y})_{i_1})_{i_2})\ldots)_{i_R} \quad . \tag{J.38}$$

Hence, the proposed procedure for the computation of the delay-free loop can be extended to any linear filter network.

J.7.3 Detection of delay-free loops

The graph proposed by Szczupak and Mitra is repeated in Figure J.8 except for the type of ordering, involving branches instead of nodes [153]. Holding the hypothesis that it represents a causal filter network (this is true, i.e., $\det \boldsymbol{F_y} \neq 0$, if, for instance, it is $b_i = i$), then ${\boldsymbol{F_y}}^{-1}$ has the structure below:

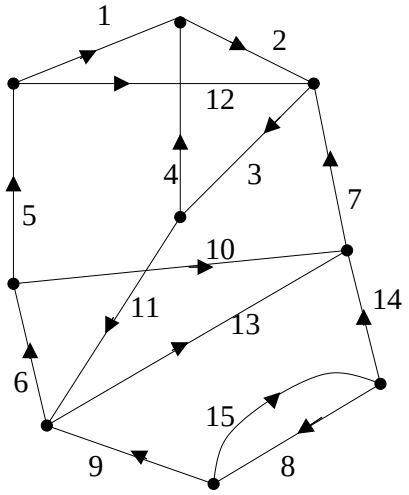

Fig. J.8. Connected graph (taken from Szczupak and Mitra, ordering branches instead of nodes).

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
2	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
3	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
4	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
5	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
6	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
7	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
8								•							○
9								•	•						•
10	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
11	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
12	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
13	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
14								•						•	•
15								○							•

where non-zero elements have been represented using the symbols '•' and '○', whereas null elements have been left blank. From that structure, using the rules given in Section J.3.1, it can be concluded that:

- branches 1-7 and 10-13 are part of the same delay-free loop (elements depicted with '•');
- branches 8 and 15 form a separate delay-free loop (elements depicted with '○');
- branches 9 and 14 do not belong to any delay-free loop (in fact they form the *oriented cut set*, alternatively figured out by Szczupak and Mitra [153]).

K

Acoustic distance for scene representation

Federico Fontana, Davide Rocchesso and Laura Ottaviani
IEEE Computer Graphics & Applications.
- SUBMITTED FOR REVIEW -

There are cases in which not all the objects and events forming the scene can be visually represented simultaneously. In those cases the scene representation can be complemented by the acoustic mode. Here we show that sounds representing objects and events in the display can be manipulated so to inform the user about her relative distance from those objects. We do this by simulating the everyday-listening experience, during which we get auditory cues from sound sources located at different relative distances from us. This kind of manipulation neither changes the nature of the sound source nor interferes with the visual information, hence its use produces an absolute increase of the bandwidth of interaction in the multimodal display. Limits in the adoption of the proposed technique come by unavoidable blurs exhibited by humans in estimating distance using the auditory system. On the other hand, representing range data through the virtual reconstruction of auditory distance cues requires neither previous training or learning effort for the user, nor the adoption of specific audio virtual reality reproduction systems.

K.1 Introduction

The representation of a virtual scene, provided to a user by a computer interface, takes advantage from the existence of information channels alternative to the visual one. Although the video plays a major role in the presentation of virtual scenarios, nevertheless the acoustic modality, in the form of speech and also non-speech output, increases the user's engagement and, provided that the audio message is correctly displayed, can enlarge the bandwidth of interaction in a multimodal display.

Improving the interactivity in a virtual environment by adding auditory information at the interface output is not straightforward, especially when non-speech communication is considered. In fact, the mere addition of sound in the display, regardless of which sounds to play, and how to display them, may result in the opposite effect, diminishing the user's performance, engagement and satisfaction.

Rather, the auditory mode requires to be correctly integrated with the rest of the display. An effective integration needs the solution of several issues of *sonification* and *spatialization* (see Sidebar 1). In the end, the user should experience a virtual environment where sounds augment the realism, providing messages that integrate the visual information and sometimes substitute it, for example when an object has been visually occluded by other elements appearing in the visual scene, or when the user is visually impaired so that her perception of a scenario strongly or completely relies on non-visual information [84].

In our research we deal with the representation of the *distance* in a user-centered (typically 3-D) virtual scenario. In principle the question is trivial: most of the objects appearing in a scene are recognized for what they are through the visual information displayed at the interface, in a way that their distance is estimated, with some approximation, as part of this information. Nevertheless, there are cases in which a multimodal, audio-visual representation of distance turns out to be useful. Given that the object/user distance emerges as an essential information in the display, those cases include (see also Figure K.1):

- objects that fall out of the visual angle. This is the case of large screens;
- objects whose relative distance the user must continuously survey, even during tasks in which the focus of visual attention is somewhere else. This happens especially during critical tasks, which prevent the user to steer the visual attention;
- objects outside the visual display. This can occur in the case of small screens;
- objects popping up somewhere, as new elements in the scene, whose existence must be rapidly declared to the user;
- objects whose relative distance can be hardly estimated by vision, unless the user starts some action such as navigating in the scene to get more information about the object: changing her viewpoint, for example, or looking for landmarks which help figuring out the object position and, hence, its relative distance. This situation can take place in virtual scenarios where the use of stereoscopic vision is not enabled, when objects which are unfamiliar to the user's eye are displayed.

In all the above cases, supporting the visual rendering model with a tool for the real-time synthesis of *auditory distance cues* can lead to advantages in the display quality, concerning the representation of the distance of objects.

K.2 Hearing distance

We want to put the user in condition to perceive distances, by integrating non-speech audio communication at the interface output. In principle, such a choice is prone to objections. In fact, even in the contexts pointed out in Section K.1 we can find alternative solutions, making use of either the visual or auditory mode.

Solutions using the visual mode may be provided, for example, by sidebars containing range indicators. In the case when the visual attention should focus on a single task, then digits informing about the distance of objects located somewhere else might be added in the visual display. Those digits should be visualized

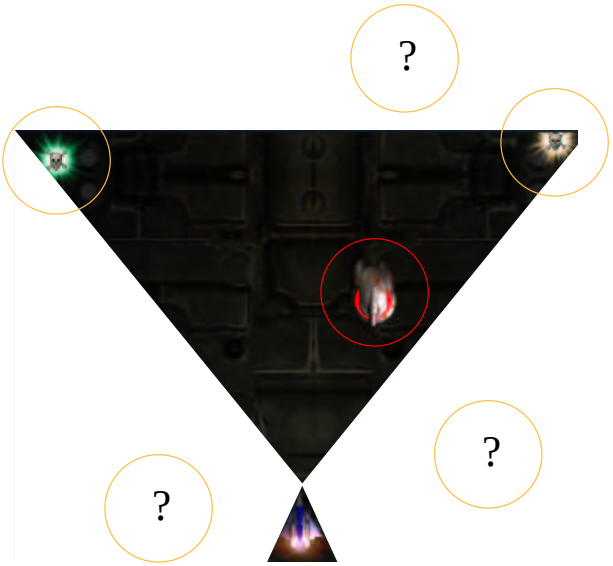

Fig. K.1. Visual displays have their own limits. Objects and/or events located outside the visual angle, outside the screen, and objects which are out of the focus of visual attention would be conveniently displayed using another modality.

close to the position in the display where the task takes place, or they could be superimposed, in the same position, as semi-transparent layered images.

The same information can be conveyed through the auditory mode, for example using a synthetic voice that tells the range to the user through periodic vocal messages or, alternatively, using non-speech sonification strategies.

Although methods which exploit either the visual or auditory mode can be worked out effectively, even leading to a precise presentation of data, nevertheless a conventional display method for presenting distance, such as those which we have just described, will probably lead to machine-to-user communication streams which saturate a non-negligible (and possibly precious) portion of the overall available bandwidth of the visual or auditory information channel, respectively.

We will show that a less conventional method can be found, for presenting the data in such a way that they fill a band of the auditory channel which is, to a wide extent, disjoined from the portion normally occupied by speech and non-speech audio communication streams. Such a band exploits a perceptual dimension that has been hardly ever considered interesting for data communication. More precisely, we are going to integrate distance *in the form of an auditory spatial cue* to existing speech or sonic messages, in a way the information on distance does not overlap with their informative content.

This strategy has a consistent background in several studies concerned with the everyday-listening experience. According to an *ecologic* approach to auditory perception, humans relate sounds to corresponding sources, from where the sounds respectively origin [61]. In this way humans perform recognition tasks, identifying each object from a corresponding sound source. Along with the recognition of an object, a localization task takes place. This task continues even after the object

recognition, if any, and especially when the object is recognized to be moving, or potentially moving.

The localization task relies on spatial cues, which are selected by our auditory system to be independent of the cues that bring information about the source. Unfortunately, *binaural listening* must deal with insuperable limits: since sound arrives at the human ears in the form of two monodimensional signals, each one entering the respective ear, then the detection of the relative position of a sound source is carried out successfully only to a certain extent. In particular, distance is one of the spatial auditory dimensions which actually suffer from non-negligible estimation blur.

Nevertheless, by most everyday-listening contexts anybody can experience that auditory distance estimation, although uncertain, is performed by listeners without any particular need of engagement, or major distraction from the focus of attention. We export this ecologic assumption in the development of a model for multimodal distance rendering using auditory cues.

K.3 Absolute cues for relative distance

Addressing auditory distance perception is a quite complicate question. The first obstacle to deal with is how to choose the experimental conditions. In fact, a trade-off must be agreed between ease of control and realism of the listening experience: the former leads in general to "poor" (sometimes unrealistic) sounds, possibly affecting the listener's judgment; the latter leads to sounds containing multiple spatial attributes, each of those potentially having a role as a conveyor of distance cues [168].

Several kinds of psychophysical experiments have been conducted, under different listening conditions:

- using real sounds, rather than loudspeaker or headphone reproduction;
- organizing setups in the open space, in the inside of various types of enclosures, in concert halls, or in *anechoic chambers*, whose extremely low wall-reflection factors enable a careful control of the experimental physical conditions, that is, precise figures of the sound pressure at the listener's ear;
- choosing between different sound sources;
- working with listeners trained at different levels.

As a result, researchers have come up with different conclusions, sometimes controversial. In some cases doubts about the quality of the equipment used in the experiments have been raised, especially for early investigations. Apart from this, evidence has been found that humans perform auditory distance estimation according to subjective perceptual scales which are generally non-linear, except for limited portions of the full range. Moreover, such scales have been discovered to be sensitive to the type of sound source used in the experiment and the characteristics of the listening environment [15].

The sensitivity to the sound source and the environment characteristics can be explained if we look at the physical, psychophysical and psychological reasons that originate it.

1. About the influence of the sound source, it must be remembered that auditory distance estimation always takes place along with a sound source recognition task. It is likely that the less familiar the source is to the listener, the more uncertain its relative distance.
   This uncertainty can be easily understood if we consider the visual counterpart. Visual perception of depth relies on precise cues of disparity existing between the two image projections contained in a stereoscopic image. Those cues are *absolute*, i.e., they do not depend on the geometric and surface characteristics of the object detected. Unfortunately, corresponding disparity cues do not exist in the acoustic modality but for sound sources located in the auditory near-field, that is, less than approximately $1 \sim 1.5$ m—indeed, in that range humans can get satisfactory visual localization cues. In other words, our auditory system cannot assess distance by comparing the two signals forming the binaural sound.
   One of the most reliable cues that in principle would enable distance evaluation is *loudness*. This cue is, to a large extent, proportional to the sound pressure measured at the ear in decibel (dB). Furthermore, the sound pressure decay in the ideal open space is equal to 6 dB for each doubling of the source/listener distance. From this, it descends that if we know the sound pressure at the source, and we are confident with the 6-dB law, then in principle we should exhibit capabilities of assessing distance by hearing. It also descends that any sound source, whose sound pressure at the emission point is unknown, cannot be located at any definite range by making use of cues of loudness. This fact is equivalently expressed saying that loudness is a *relative* cue for distance evaluation.
   In practice, experiments aiming at testing distance estimation for known sound sources in conditions of 6-dB decay have shown that even well-known sources, such as voice, lead to estimation mismatch and sensitivity to the voice intonation. This conclusion is consistent with the unfamiliarity people have with the ideal open-space. In fact this environment can be experienced, with some approximation, in contexts which are quite special for most of the people, such as large open fields covered with snow.
2. About the influence of the listening environment, it is true that any enclosure adds unique attributes to a sound during its journey from the point of emission to the listening position. Indeed, those attributes vary both with the emission and listening point. They appear in the form of *reverberant energy* in the signal immediately following the direct sound (see Sidebar 2). Once again, this reverberant energy does not translate into binaural cues (though, from a purely physical viewpoint, it does). Furthermore, reverberation cues are less reliable than loudness cues, especially if the subject has not been given a prior overview of the room size, volume and overall characteristics.
   Nevertheless, reverberant energy defines *absolute* cues. For proving that, we can just play a sound source in a small room made of concrete walls, then play the same sound source in the inside of a concert hall having walls covered with acoustic fabric. In the latter case, the source will sound "warmer" and more "enveloping". On the other hand, we will probably not be able to go in much more detailed description of our auditory impression.

> Experiments show that humans make use of reverberation cues during distance estimation tasks [66]. Reverberant energy, although less reliable than loudness, gives precious absolute information that helps subjects to assess distance especially in the case when they have low or no familiarity with the sound source. Moreover, reverberation is part of our everyday-listening experience: although during the day we move around many enclosures whose reverberant characteristics differ, sometimes noticeably, nevertheless our expectation of their gross reverberant characteristics (those we can actually perceive) is to a good extent acquired from experience.

We conclude that we can reasonably expect to render distance by adopting a model that adds reverberant energy to sounds to an amount which will be proportional to the source/listener distance, provided that it also varies the loudness of the sound source accordingly.

Returning for a moment back to the question about the visual and auditory bandwidth occupation, now we can figure out that such a model will neither interfere with any type of acoustic message bringing information about the sound source characteristics, nor with any sonification paradigm. Of course, it will not occupy bandwidth pertaining to the visual mode. Furthermore, the information it brings will not distract the user from her main activity. In other words, we expect that using this model will increase the overall informative bandwidth the user can get from the multimodal display.

On the other hand, such a model is also expected to lead to subjective interpretations of the auditory distance cues. As we have seen, this limit must be to a large extent reconducted to unavoidable idiosyncrasies existing in humans' auditory distance perception rather than to limits due to the model itself. This drawback must be accepted if we want to pick up the potential advantages of auditory distance rendering. Otherwise, in the case when exact distance information is needed, even for objects and events which are outside the focus of visual attention, then the interface designer should steer to alternative ways of presenting those data.

K.4 Displaying auditory distance

The systems devoted to add spatial cues to sounds in the form of reverberant energy are called *reverberators*. Typical fields of application for those systems are in the industry of music and entertainment. Reverberators usually operate in the digital domain, and filter the sound according to processing strategies which are often kept undisclosed to the user. This must not surprise, since the search for an optimal reverberation model is mainly a matter of proprietary "soundcrafting" activity [58].

The use of reverberation cues as information conveyors in a human-computer interface is perhaps a novel issue. Commercial reverberators do not use to implement distance as a control parameter: this sounds quite natural, since their scope does not include applications of multimodal display. Rather, they aim at rendering other aspects of reverberant sounds, such as their *warmth*, *envelopment*, and other cues that certainly pertain to the musical field, but are fairly interesting in human-computer interaction. For these reasons, the design of a reverberator

follows criteria that are strictly oriented to the achievement of a pleasing rather than informative sound rendering.

Our model, on the other hand, is strictly devoted to distance rendering. Our design criteria are driven by three requirements:

1. matching the psychophysical evaluation;
2. providing easy and direct access to the driving parameters, both at the early design step and in the final application;
3. having low computational impact, this being a "must" for any system which is supposed to work in real-time.

Moreover, we cannot forget that the model is supposed to evolve into an application whose execution takes place in systems, which are often located in physical environments where groups of people actively work together, possibly joining cooperative tasks. These contexts are not so obvious to deal with when the auditory mode comes into play, since acoustic messages displayed by interfaces for individual personal spaces must coexist in the same working environment, avoiding mutual interference.

In that situation the use of audio headphones solves the problem only partially, since this choice prevents people from using direct communication. One possible solution comes from the adoption of *open* headphones, or, if the working conditions permit that, the adoption of low-power "personal" loudspeaker systems.

Both solutions in principle provide a sufficient quality of the audio message, although the transmission to a listener of cues of audio virtual reality requires the use of more sophisticate methods and equipment, which are still confined to specific (and expensive) VR installations devoted to particular purposes, and to the research laboratory [57]. In this sense researchers and developers are producing a special effort to export audio VR to less specific contexts.

On the other hand, auditory distance cues have been proved to be quite robust with respect to the type of equipment chosen for the presentation [168]. This robustness is a key point for the development of our model, since it prevents (at least at an early development stage) from necessarily using an audio VR display system as presentation tool.

An exception occurs when the source to be acoustically rendered results in too low sound pressure values at the listener's position. This occurs for example when a normally talking person is located in the far field (say, around 10 m) or, equivalently, when a weak sound is emitted in the medium field. In that case, the message displayed to the user via open headphones would be so soft that it could hardly be heard, especially in presence of external noise. For this reason, we chose to emphasize reverberation instead of loudness distance cues. Although this choice produces exaggerate (but not unpleasant) reverberation, nevertheless we expect that those cues are more robust to external noise and far-field distance rendering compared with loudness, meanwhile still being informative about distance. Indeed, increasing loudness with an aim to emphasize sources located in the far field distorts the information at least as much as exaggerating reverberation. Furthermore, loudness changes (even those set on the fly by the user) devoted to counterbalance external noise result not only in information distortion, but also in a difficulty of the user to cooperate with the rest of the group due to excessive loudness of her

audio, if not in an overall performance degradation of the workgroup activity if several audio messages coming from different working positions, displayed through loudspeakers, start to interfere and mix together.

K.5 Modeling the virtual listening environment

Consider a child playing inside one of those tubes that are found in kindergartens. If we listen to the child by staying at one edge of the tube, we have the feeling that he is located somewhere within the tube, but his apparent position turns out to be heavily affected by the acoustics of the tube. Nevertheless, as we move back and forth along the tube we can experience consistent changes in the auditory cues, in a way that we can perceive whether we are approaching the child, or moving away from him.

This metaphor is the basic building block of our interface. Consider a listening environment made of a tube, and imagine to put a sound source at one end of it. Then enable a listener to move inside the tube along its main direction, in a way that she can record in some specific points the sounds coming from that source. Collect those sounds to form a set of audio samples, where each sample will finally "label" a corresponding distance.

Note that the tube is not supposed to physically contain the sound source and the listener: since all the actors in this metaphor are virtual, we can think of the source and the listener just as small (or even point-wise) emission and receiving points, respectively; hence, "shape" the tube, in particular its section, to a size which fits most with our main requirement, that is, conveying distance cues. This kind of abstraction has a solid tradition in the modeling of listening environments where the sound sources are not physically present, but reproduced using loudspeakers.

We experimented with several tube sizes and configurations, until we found a virtual tube that seemed to be effective for distance rendering. In this research we were helped by a virtual acoustics tool that allows to set up resonant cavities having the desired requirements of size and geometry, along with specific parameters of wall absorption and wall reflection exhibited by the internal surfaces of the tube (see Figure K.2).

Finally, we came up with a square-sectioned cavity sized $9.5 \times 0.45 \times 0.45$ m modeling the tube depicted in Figure K.3. The internal surfaces of the tube were designed to exhibit natural absorption properties against the incident sound pressure waves. The surfaces located at the two far edges were modeled to behave as *total* absorbers, in a way that they do not reflect any incident sound wave back to the inside of the tube.

The possibility to choose the geometry of the cavity, along with the local control over the properties of its surfaces, gives high flexibility and freedom of choice in modeling our listening environment, probably more than what is actually needed during the design of a reverberator.

Such a flexibility was made possible by adopting a particular numerical scheme for simulating the propagation of pressure waves along the cavity. This scheme, called Waveguide Mesh (WM), enables the designer of virtual acoustic environments to manipulate the shape of the cavity and the main properties of its surfaces

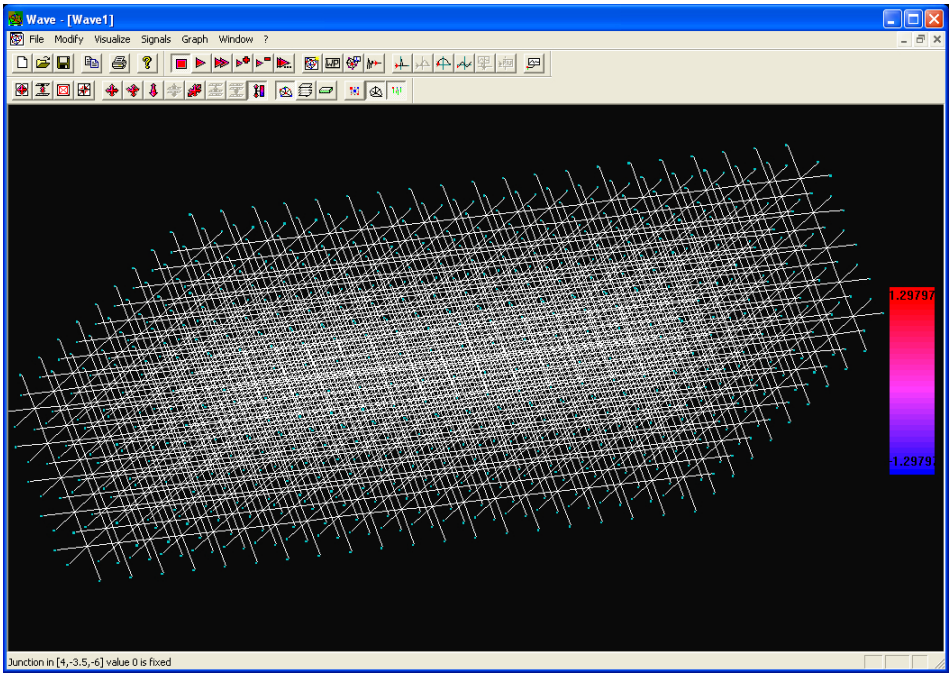

Fig. K.2. Using the above desktop application, we could experience several configurations of our tubular model by changing its size and geometry, along with the parameters of wall absorption and wall reflection exhibited by its internal surfaces.

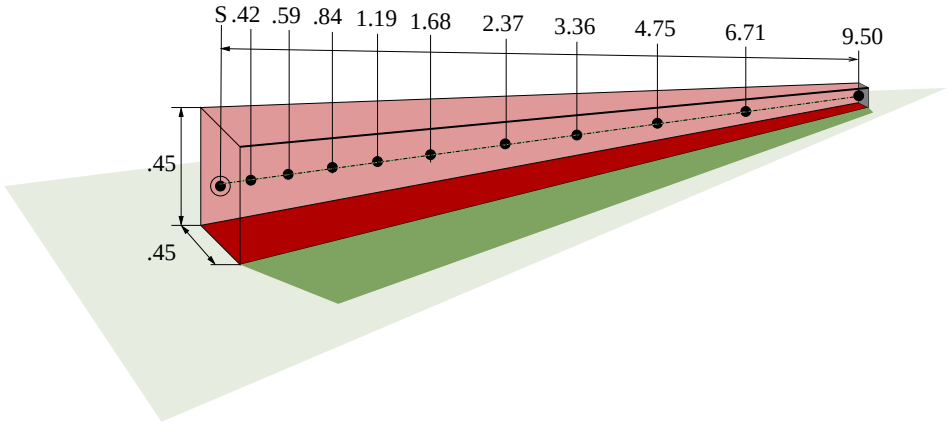

Fig. K.3. The virtual acoustic tube. All distances from the sound source (S) expressed in meters.

directly, obtaining the results from the simulation in a reasonable time even using a desktop application such as the one in Figure K.2, which was not optimized for speeding up the execution of the algorithm but, rather, for observing the evolution in time of the sound waves along the propagation domain.

Further details on the WM are addressed in Sidebar 3. In our model, each couple of adjacent nodes forming the mesh are separated by a distance (or *spatial step*) d, which is in relation with the sound speed c and the sampling frequency F_s chosen for the model via the following formula:

$$d \geq \sqrt{3}\,\frac{c}{F_s} \quad . \tag{K.1}$$

This formula accounts in particular for the numerical error introduced by the WM, known as *dispersion*, which causes some components of the acoustic wave to travel slower along the mesh. In other words, d is a function of those components—if all of them traveled at the same speed, as it happens in the case of sound propagation along an ideal medium, then we should change the symbol '$\geq$' into '$=$'. For this reason we have to keep in mind that certain high frequency traveling components must "face" larger spatial steps (up to around 20%) compared to low frequency traveling components, for which the symbol of equality holds with good precision.

We can reasonably expect that dispersion influences the experiment only to a negligible extent. Though, a preliminary test aiming at assessing the humans' capability to perceive dispersion, such as that affecting our WM, should be conducted for a rigorous validation of the model.

Assuming the speed of sound to be equal to the value taken in air in normal humidity conditions, i.e., $c = 343$ m/s, and once dispersion has been neglected, then we come up with a spatial step equal to $d = 74.3$ mm, given that the model is executed at a frequency $F_s = 8$ kHz. Finally, our tube model requires $127 \times 5 \times 5 = 3175$ nodes, each one of those runs 11 adds and 1 multiply at any time step.

The sampling frequency we have chosen is fairly low. In particular, this frequency enables the model to process sounds whose spectral energy does not exceed the Nyquist value $F_s/2 = 4$ kHz. This choice was motivated by the need of testing the performance of a realization representing a "lower bound" in our model quality.

Besides that, Equation (K.1) implies that the number of nodes in the mesh increases with the sampling frequency according to a cubic progression. Moreover, increasing F_s also means that a proportionally higher number of sound samples must be processed. Essentially, the number of operations required by the model to process an input sound depends on the sampling frequency by a factor 10^4.

WMs enable to apply *digital waveguide filters* (DWF) at the mesh boundary (again, see Sidebar 3 for further details on the DWF model). Those filters can be easily tuned to simulate local reflection and absorption properties of the internal surfaces of the tube, which are modeled in correspondence of the mesh boundaries. Our DWF realizations simulate 1st-order spring/damper systems, such as the one depicted in Figure K.4, modeling simplified resonating/absorption properties of the wall surfaces.

Using the simple physical system just described, the resulting DWF terminations are made of 1st-order pole-zero filters. Despite their simplicity, those filters can be tuned to closely approach the reflective properties of real absorbing walls. Considering that the surfaces at the two terminations of the tube have been set to be totally absorbing (recalling Figure K.4, this means that for those surfaces it is $p_r \equiv 0$), then the total number of boundary DWFs results to be equal to

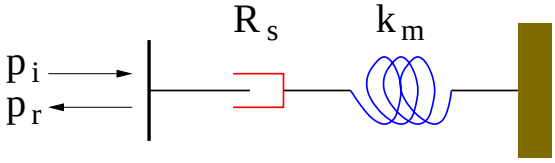

Fig. K.4. Local wall reaction to incoming sound pressure waves p_i is modeled by 1st-order DWFs, simulating simple damper/spring systems characterized by having damping parameter R_s and spring constant k_m. Those parameters tune the absorption and resonance characteristics of the wall surfaces, respectively.

$127 \times 5 \times 4 = 2540$. Each one of those filters needs 2 sums and 3 multiplies to perform its computation at any time step.

K.6 Model performance

We now investigate whether a configuration of our tubular virtual environment exists, such that the model meets all the requirements and expectations addressed in the beginning. Prior to the psychophysical experiment we measured some figures that are closely related with the reverberant energy produced by the model.

We put a sound source at one end of the tube (labeled with S in Figure K.3) along the main axis. Starting from the other end, we moved a listening point along 10 positions $x_{10}, \ldots, x_1$ over the main axis, in such a way that, for each step, the source/listener distance was reduced by a factor $\sqrt{2}$. Finally the following set X of distances (expressed in meters) comes out, as shown also in Figure K.3:

$$X = \{x_i,\ i = 1, \ldots, 10\} = \{0.42, 0.59, 0.84, 1.19, \ldots, 4.75, 6.71, 9.5\} \qquad \text{(K.2)}$$

Ten stereophonic impulse responses measuring 0.25 s (i.e., 2000 signal samples at 8 kHz) were acquired from the tube model along positions $x_1, \ldots, x_{10}$. The right channel accounted for acquisition points exactly standing on the main axis, whereas the left channel accounted for points displaced two nodes away from that axis, this corresponding to an interaural distance of about 15 cm.

A plot of the audio file derived by queuing those responses according to the sequence $\{x_{10}, \ldots, x_1\}$ is given in Figure K.5. From a rapid analysis of this plot, especially in the left (upper) channel, it comes out that the energy of the direct signal, corresponding to the magnitude of the first part of each response, increases for decreasing distance. Conversely, the *reverberation time*, related with the time length of each response, increases with distance.

Using these responses, the sound is heard by a listener to come slightly from the left. Consistently with the geometry of the virtual environment, this sensation is very weak for listening points located in the far-field, but it increases as long as the virtual listening position is moved closer to the sound source. We decided to experiment using stereophonic samples where the two channels convey different sounds, such as those produced using the responses plotted in Figure K.5, in order to minimize effects of *inside-the-head* localization typically arising when monophonic sound material is auditioned through conventional headphone equipment, without prior specific binaural processing [15]

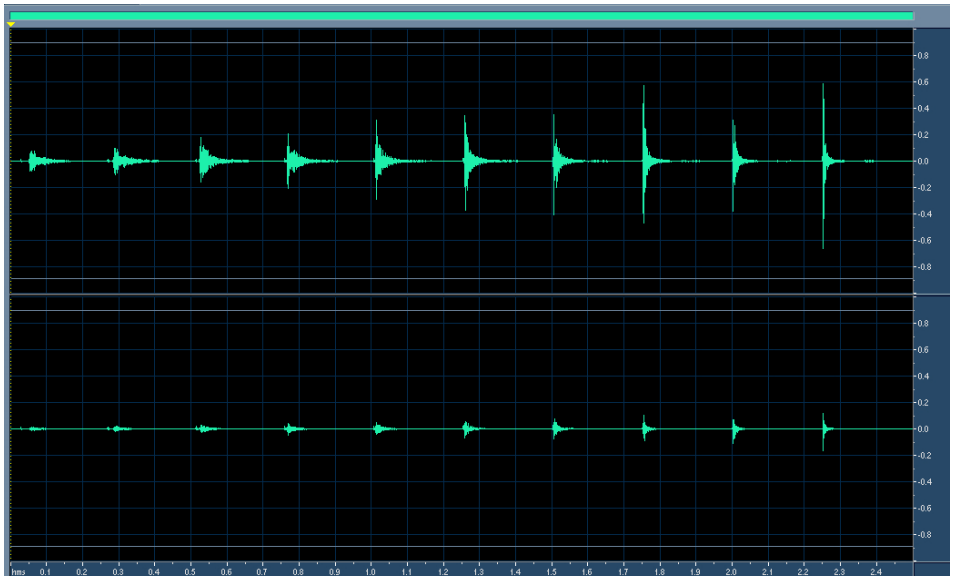

Fig. K.5. Plot of the audio file containing all the impulse responses computed by the tubular model, according to the sequence $\{x_{10}, \ldots, x_1\}$.

We then convolved these responses with a monophonic, short anechoic sample of a cowbell sound, and labeled the resulting 20 output sounds according to the indexes and channels of the respective impulse responses: $1L, \ldots, 10L$, and $1R, \ldots, 10R$. So, if we label each response, say h, with its own index, and denote with x the input sound, then, using the same notations, we have that the outputs are equal to

$$y_i = x * h_i \quad , \quad i = 1L, \ldots, 10L, 1R, \ldots, 10R \quad , \tag{K.3}$$

respectively. The symbol '$*$' denotes convolution.

The unfamiliarity of the sound source to the listener was considered an important requirement for the experiment, in a way that subjects could not rely on absolute loudness cues to assess the source/listener distance. Though, the use of an unrealistic (hence, non-ecologic) sound was avoided.

Figure K.6 gives some figures about the samples used in the experiment. On the above side we can see average output magnitudes in dB, calculated by

$$10 \log \frac{1}{n} \sum_n [y_i(n)]^2 \quad , \quad i = 1L, \ldots, 10L, 1R, \ldots, 10R \quad , \tag{K.4}$$

as functions of logarithmic distance. It appears that, for distance increasing, those magnitudes decrease (also in average) more slowly than the 6-dB law, which has been added in the plot in dashed line. This is actually what we first demand to the model. Furthermore, on the below, spectra calculated from the outputs $1L$, $1R$, $10L$, $10R$ have been plotted. They confirm that each output exhibits its own definite character, that in principle enables the user to discriminate between individual distances using cues different from loudness.

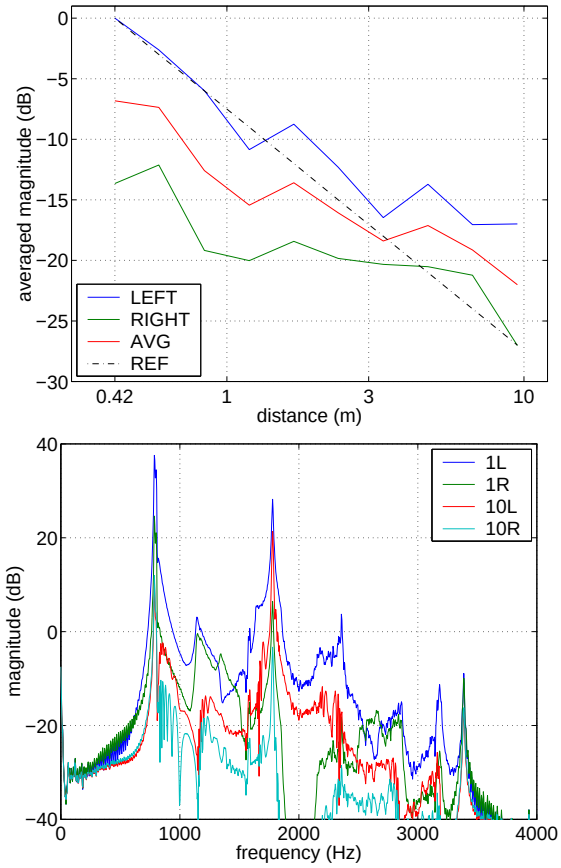

Fig. K.6. Average magnitude plots of outputs $1R, \ldots, 10R, 1L, \ldots, 10L$ plus their averages as functions of logarithmic distance, along with reference 6-dB law (above). Magnitude spectra of outputs $1L$, $1R$, $10L$, $10R$ (below).

K.7 Psychophysical evaluation

We have conducted two experiments in two different listening conditions: one using headphones normally available in the market, the other using high-quality loudspeakers. As mentioned before, we were interested neither in adopting any specific presentation strategy, nor any virtual audio technique aiming at optimizing the sound to the headphone or loudspeaker reproduction.

The listening tests are based on the magnitude estimation method without modulus [45], by means of which we investigated how subjects scaled the perceived distance and, hence, whether our model is effective or not.

The setup involved a PC Pentium III, with a Creative SoundBlaster *Live!* soundcard. During the first experiment sounds were auditioned through Beyerdynamic DT 770 closed headphones. During the second experiment, the participants sat 1.5 away from a pair of Genelec *2029B* stereo loudspeakers, 1 m far from each other, and a Genelec subwoofer located in between the loudspeakers.

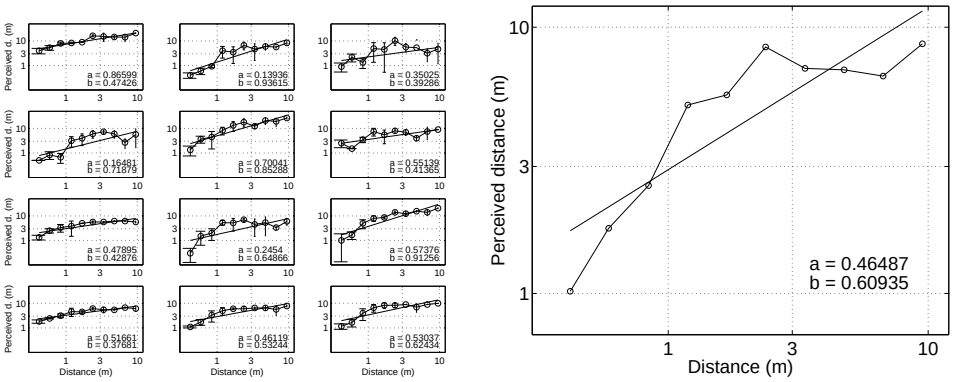

Fig. K.7. Headphone listening: Individual distance evaluations together with individual linear regression lines. a: intercept. b: slope (left). Average distance evaluation together with linear regression line. a: intercept. b: slope (right).

K.7.1 Headphone listening

12 subjects, 4 female and 8 male students and workers at the University of Verona with age between 22 and 40, voluntarily participated to the experiment involving headphone reproduction. All of them were naive listeners.

In a quite (but not silent) office room we presented to each subject a sequence of 30 sounds, randomly repeating for three times the 10 sounds taken from the set of stimuli. For each sound sample, subjects had to estimate how far they were from the sound source, reporting a value accounting for distance (in meters, either integer or decimal) after listening to each sample.

The experiment was conducted without training. After rating the first sample, participants attributed values to the following samples proportionally to the first estimation. Moreover, since we did not set a modulus, the collected data defined scales depending on the individual listeners' judgments. These scales range from 0.2-8 m (subject no. 8) to 1-30 m (subject no. 5).

The three judgments were then geometrically averaged for each subject. Then, the resulting average values were used to calculate a mean average. Finally, we obtained a common logarithmic reference scaling subtracting the mean average from the individual averages.

In fig. K.7 (left) the distance evaluations as functions of the source/listener distance are plotted for each subject, together with the corresponding linear functions obtained by linear regression. The average slope is 0.6093 (standard deviation 0.2062), while the average intercept is 0.4649 (standard deviation 0.2132).

In fig. K.7 (right) the perceived distance averaged across values is plotted as function of the source/listener distance, together with the relative regression line ($r^2 = 0.7636$, $F(1,8) = 25.8385$, $F_{\text{crit}}(1,8) = 11.2586$, $\alpha = 0.01$). The r^2 coefficient is significant at $\alpha = 0.01$ and, therefore, the regression line fits well with the subjects' evaluations.

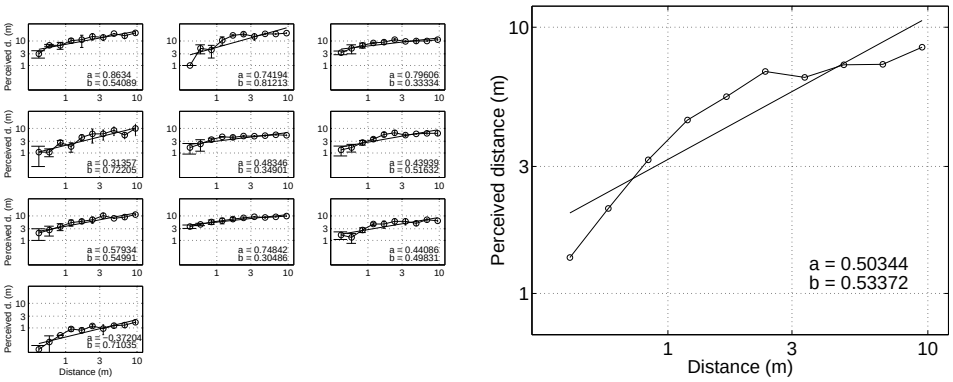

Fig. K.8. Loudspeaker listening: Individual distance evaluations together with individual linear regression lines. a: intercept. b: slope (left). Average distance evaluation together with linear regression line. a: intercept. b: slope (right).

K.7.2 Loudspeaker listening

The second experiment involved 4 female and 6 male volunteers aged between 23 and 32, who worked or studied in our department. Four of them had participated to the first experiment as well.

The set of stimuli was the same as for the previous test. Listeners joined the same office room, and were blindfolded in order to minimize the influence of factors external to the experiment. The test was performed exactly in the same way as before except that, from the listening point, they communicated the rated distance value to the experimenter, who wrote down the data. As in the previous test, the first value determined the subjective scale.

The same figures obtained during the experiment with headphones are repeated in Figure K.8 concerning the second experiment. This time the average slope is equal to 0.5337 (standard deviation 0.1741) and the average intercept is equal to 0.5034 (standard deviation 0.3573). Furthermore we have $r^2 = 0.8512$, $F(1,8) = 45.7603$, $F_{\text{crit}}(1,8) = 11.2586$, and $\alpha = 0.01$. The r^2 coefficient is significant at $\alpha = 0.01$ so, once again, the regression line fits well with the subjects' evaluations.

K.7.3 Discussion

From both experiments we observe that subjects overestimate the distance for sound sources that are close to the listener, and that they reduce this overestimation for greater distances. This result is interesting since it partially conflicts with some recent experiments conducted in real listening environment. In particular, Zahorik [168] reports the tendency of listeners to overestimate near-field, and underestimate far-field sound source. Furthermore, compared to his experiment our point of correct estimation is located farther.

This evidence probably follows from the exaggerated reverberant cues produced by our model. This conclusion does not contradict our initial aim, i.e., synthesizing distance cues which are strong enough to lead listeners to rate the perceived sounds as if they came from sources located in the far field.

One subject (no. 10) in the second experiment perceived all the sound sources to be closer compared with the first experiment. However, during the talk/questionnaire following the listening session this participant did not reported any difficulty in performing the required task.

Furthermore, there is no evident difference between judgments of naive participants, and subjects trained by the previous experiment.

K.8 Conclusion and future directions

The proposed model matches, at least to some extent, three important issues that must be kept in mind during the development of a model for the display of data to the user:

- **Consistency with the human interface.** We tried to optimize from a psychophysical viewpoint the presentation to the user of distance data, in a way that he does not have to produce any special effort to acquire them, despite some lack of precision in the acquisition due to inherent limits in humans' auditory evaluation of distance.
- **Versatility.** The tuning parameters of the tubular model are easily accessible, along with its input and output point. Thus, different distances can be immediately selected along with specific tunings of the model itself. Furthermore, effective information is conveyed to the user avoiding the adoption of peculiar systems for audio VR, most of which need subjective tuning. Also, an effort has been made to minimize the impact of the model when it must be employed in cooperative environments, despite unavoidable application limits shown by audio interfaces in these contexts.
- **Technological impact.** The tubular model has been implemented off-line. This means that a desired set of impulse responses can be calculated from the model, in such a a way that they can be later convolved in real time with any sound source, possibly free of reverberant energy, to add spatial cues accounting for distance. The time length of such responses is not critical. Thus, the dimension of the set of responses can be limited to a reasonable size. Furthermore, there are different ways to keep this set compact, either using audio coding strategies or taking advantage from specific interpolation techniques. Hence, even a fine-grained distance rendering can be obtained using the tubular model, providing the audio interface with a digital convolver such as those which are actually available in the market. Finally, conventional digital audio processing techniques can be applied to the convolved sounds, to add binaural auditory cues accounting for the relative angular position, or azimuth, of the sound source [21].

One of the recent research projects activated in our Video Image Processing and Sound (VIPS) lab deals with the design of a navigation aid for visually impaired people, provided with a machine interface having video input mode and audio output mode. This application includes a pair of video cameras, by which sequences of stereoscopic images are captured from the visual scenario. Those images are then processed by an algorithm of visual analysis, to recognize objects in the scene. Such

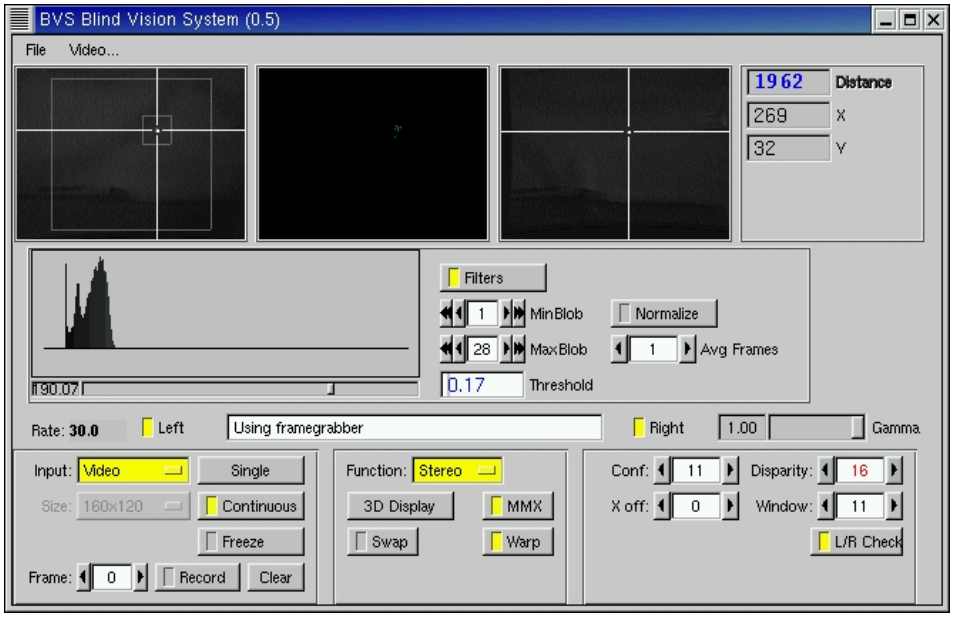

Fig. K.9. Control panel of the application for navigation aids developed at VIPS.

objects are finally presented at the audio interface: they are sonified according to their macroscopic geometric features, and spatialized according to their angular position and distance relative to the user.

The distance is presented to the user through the audio channel, by convolving the sounds accounting for the object geometries with pre-recorded responses calculated using the virtual tube. A snapshot of the control panel used for tuning the system during its development is shown in Figure K.9.

This navigation aid is now going to be tested on a palmtop portable device, whose display has been substituted with the audio-video interface. After succeeding with this porting we expect to start experimenting with a group of blind persons, with an aim to test the interface and check whether those persons can take advantage from augmenting their auditory and haptic representation of the surrounding environment with additional auditory cues, such as those provided by the system.

Looking forward to a virtual tube running in real-time, the figures about DW nodes and DWFs we gave in Section K.5 imply that a real-time application running the virtual tube at 8 kHz would require about 340 MIPS in a typical DSP implementation. The cost of such a computational power does not seem to be justified for an application like this. On the other hand, the continuously increasing power of real-time devices, and the inherent affinity of the WM with distributed, modular architectures, allows to envision a real-time inexpensive implementation of the virtual tube in the future.

This research came out as part of Dr. Fontana's and Dr. Ottaviani's PhD activity made in the Department of Computer Science at the University of Verona. Dr. Fontana's dissertation is available on the web, along with audio and video sam-

ples which illustrate the use of the virtual tube together with different types of sound sources and sonification paradigms, and inside the navigation aid developed at VIPS. Those samples can be found at `http://profs.sci.univr.it/~fontana/`.

K.9 SIDEBAR 1 – Sonification and Spatialization: basic notions

1. *Sonification* looks for effective rules that associate displayed objects with sounds, in a way that the auditory part of the display efficiently conveys informative and discriminable messages about specific properties of those objects [84].

2. Sound *spatialization* deals with the localization and characterization of sounds in a three-dimensional scenario. As a branch of psychophysics it is concerned with the auditory cues, by which we (also, but not only) attribute a position, relative to the listener, to the perceived sound sources. As a branch of sound processing it provides models that add specific attributes to "raw" (more precisely, *anechoic*) sounds, in a way that those sounds are finally characterized by containing spatial auditory cues. Those models are usually implemented by digital sound post-processing devices.

The most significant achievement obtained in the field of auditory spatialization has been understanding to a very large extent

- how humans perceive the relative angular position of a sound source, in terms of its *azimuth* and *elevation*;
- how to quantify the cues responsible for the perception of the source angular position, by means of the *Head Related Transfer Functions (HRTF)* model;
- finally, how to re-synthesize HRTFs to subjectively reproduce acoustic virtual scenarios, containing sound sources characterized by having a definite angular localization [15].

K.10 SIDEBAR 2 – Characteristics of Reverberation

Transmission of sound between a source and a listener is affected by reverberation as soon as the acoustic waves encounter, during their travel from the source to the listening point, *reflective* bodies [58]. In our everyday-listening experience we cannot hear sounds that are free of reverberant energy. In fact, only a few materials absorb the acoustic waves to such an extent that the energy reflected from such materials cannot be heard.

An exception comes from the *anechoic* listening environments. Due to the application of special arrangements of damping materials over the room walls, and the absence of reflective objects inside the room, anechoic chambers define listening environments that, although artificial, realize with good approximation the ideal of anechoic listening. Anechoic chambers become necessary as long as experiments must be performed in controlled acoustic conditions.

Figure K.10, on the above, shows a common everyday-listening context. The telephone ringing is heard by a listener first by the direct sound 1, whose path is

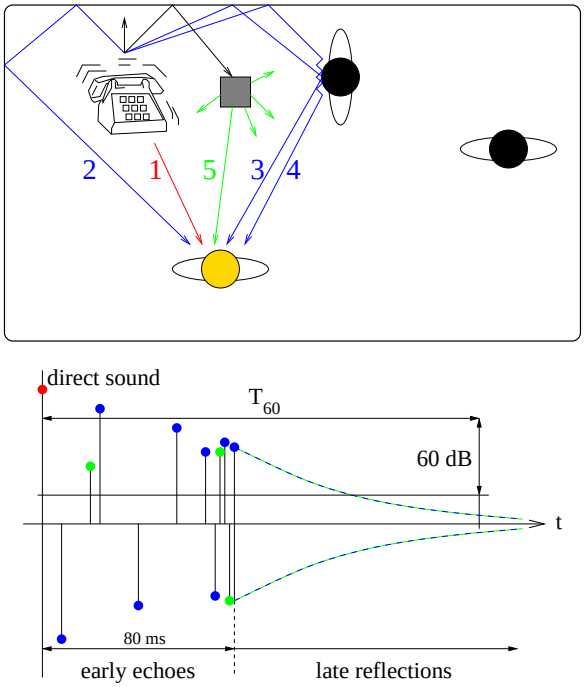

Fig. K.10. Reverberation. Direct and indirect acoustic wave transmission (paths 1 and 5, respectively) and reflection (paths 2, 3 and 4) (above). Qualitative plot of a reverberant acoustic signal (below).

highlighted by a red arrow. Since the telephone emits its sound around all directions, then many echoes reach the listener after the direct sound. The characteristic of each echo depends on the nature of the bodies the corresponding acoustic wave has encountered during its travel.

As a rule of thumb, walls provide dry, low-scattered reflections (the blue path, number 2, shows an example of wall reflection). Objects and furniture located inside the enclosures define softer, more diffuse wave reflection and transmission (green arrows; path number 5 shows an example of wave transmission). Other persons in the listening environment reflect sounds as well (paths 3 and 4).

In consequence of reverberation, the listener receives multiple echoes. Figure K.10, on the below, shows a qualitative plot of the signal as it is received at the listening point. This signal is made of the direct sound, plus reverberant energy in the form of multiple echoes.

The so-called *early echoes*, occurring within around the first 80 ms after the direct wave, introduce specific timbral characteristics in the sound. Early echoes are still distinguishable in a sound sample measured using conventional acoustic equipment.

The number of echoes received at the listening point increases with time. As soon as their density reaches a sufficiently high value, *late reflections* can be satisfactorily characterized by a statistical description, according to which the number of echoes N_e occurring after a time t depends only on the volume V of the room:

$$N_e(t) = \frac{K_e}{V} t^3 \quad , \tag{K.5}$$

where K_e can be considered constant in most everyday-listening situations.

On the other hand, the energy of those reflections decreases with time. The resulting effect on a dissipative room is an exponential decay of the sound energy that, once again, can be statistically described using macroscopic parameters such as room volume and active absorbing surface area, S_α. A simplified version of the *Sabine's formula* calculates the 60-*dB reverberation time* T_{60} (plotted also in Figure K.10) as

$$T_{60} = 0.163 \frac{V}{S_\alpha} \quad . \tag{K.6}$$

The spectral characteristics of reverberant sounds are complicate to describe with precision as well. In fact, the number of resonances N_r introduced by reverberation in the sound increases, with frequency f, to a point above which it can be statistically calculated to be equal to

$$N_r(f) = K_r V f^3 \quad , \tag{K.7}$$

where, again, K_r is constant in most everyday-listening conditions.

So, models aiming at reproducing reverberation must rely on methods which simulate the temporal and spectral characteristics of reverb only on a perceptual basis. Otherwise, the requirement of real-time could not be observed in the application. Nevertheless, successful artificial reverberators have been designed for the music and entertainment market.

K.11 SIDEBAR 3 – The Waveguide Mesh

WMs define ideal numerical schemes for the simulation of acoustic linear pressure wave fields, characterized by having constant parameters along the propagation domain. Moreover, WM boundaries model locally reacting surfaces successfully.

The core feature of this scheme is its capability to process acoustic *waves* instead of acoustic signals [13]. One-dimensional pressure waves are related to the corresponding pressure signal through the following fundamental wave-based decomposition:

$$p = p^{(i)} + p^{(r)} \quad , \tag{K.8}$$

where p is the pressure signal, and $p^{(i)}$ and $p^{(r)}$ are conventional *incident* and *reflected* pressure waves, respectively.

WMs result from putting together modular elements, called *lossless scattering junctions*, but renamed in this paper as *nodes*, to form a 3-D regular grid. Although several grid configurations are possible, for the virtual tube we have used the original orthogonal configuration since it allows straightforward implementation and analysis.

According to this configuration, each node is enabled to receive six incident waves coming from corresponding adjacent nodes. From that, each node then instantaneously scatters out six output waves according to the following formula:

$$p_l^{(r)} = \frac{1}{3}\sum_{k=1}^{6} p_k^{(i)} - p_l^{(i)} \quad , \quad l = 1, \ldots, 6 \quad , \tag{K.9}$$

where $p_l^{(i)}$, $l = 1, \ldots, 6$, and $p_l^{(r)}$, $l = 1, \ldots, 6$ are the six pressure waves incident and scattered out to and from the node, respectively. Lossless scattering junctions model ideal scattering of pressure waves occurring at the intersection point of 6 air columns joint together. Also, the pressure signal over this point can be immediately calculated:

$$p = \frac{1}{3}\sum_{k=1}^{6} p_k^{(i)} = \frac{1}{3}\sum_{k=1}^{6} p_k^{(r)} \quad . \tag{K.10}$$

Adjacent nodes communicate by exchanging incident and scattered waves mutually. The explicit computability of the scheme is guaranteed by the *delay-pass* rule, according to which waves scattered at one time step toward an adjacent node become incident waves to that node at the next time step. This rule also determines the speed of each component forming a traveling wave, which can be immediately figured out by expliciting c in (K.1).

Figure K.11 depicts a particular of the WM model of the virtual tube. In particular, WM boundaries are modeled by DWF terminations. Similarly to the nodes, those filters acquire incident waves, and produce reflected waves instantaneously according to the filter coefficient. Their physical meaning is quite clear, in fact they explicitly simulate wave reflection of pressure waves against a surface, whose properties are transferred into values of DWF coefficients.

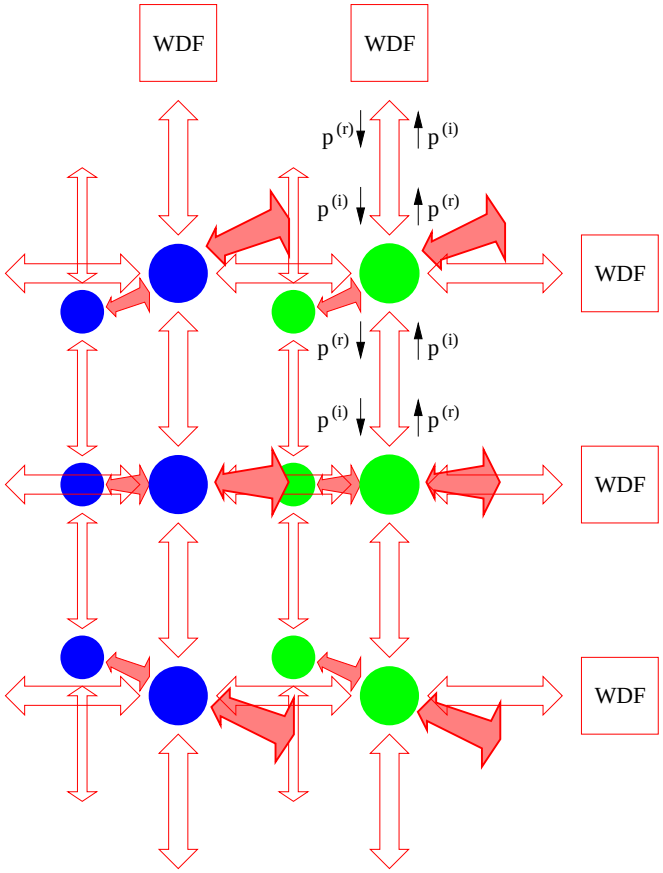

Fig. K.11. Particular of the 3-D orthogonal WM. Nodes belonging to different vertical planes (i.e., orthogonal to the plane defined by the sheet) have been drawn using different colors. DWFs modeling two vertical boundary surfaces are also visible.

References

1. J. M. Adrien. The missing link: Modal synthesis. In G. De Poli, A. Piccialli, and C. Roads, editors, *Representation of Musical Signals*, pages 269–298. MIT Press, Cambridge (MA), 1991.
2. M. Aird, J. Laird, and J. ffitch. Modelling a drum by interfacing 2-d and 3-d waveguide meshes. In *Proc. Int. Computer Music Conf.*, pages 82–85, Berlin, Germany, August 2000. ICMA.
3. V. R. Algazi, C. Avendano, and R. O. Duda. Elevation localization and head-related transfer function analysis at low frequencies. *J. of the Acoustical Society of America*, 109(3):1110–1122, March 2001.
4. V. R. Algazi, R. O. Duda, R. P. Morrison, and D. M. Thompson. Structural composition and decomposition of hrtfs. In *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pages 103–106, Mohonk, NY, October 2001. IEEE.
5. J. B. Allen and D. A. Berkley. Image method for efficiently simulating small-room acoustics. *J. of the Acoustical Society of America*, 65(4):912–915, April 1979.
6. F. Avanzini and D. Rocchesso. Controlling material properties in physical models of sounding objects. In *Proc. Int. Computer Music Conf.*, pages 91–94, La Habana, Cuba, September 2001.
7. F. Avanzini and D. Rocchesso. Modeling collision sounds: Non-linear contact force. In *Proc. Conf. on Digital Audio Effects (DAFX-01)*, pages 61–66, Limerick, Ireland, December 2001.
8. C. Avendano, V. R. Algazi, and R. O. Duda. A head-and-torso model for low-frequency binaural elevation effects. In *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pages 179–182, Mohonk, NY, October 1999. IEEE.
9. D. R. Begault. *3-D Sound for Virtual Reality and Multimedia*. Academic Press, Boston, MA, 1994.
10. L. L. Beranek. Concert hall acoustics - 1992. *J. of the Acoustical Society of America*, 92(1):1–39, July 1992.
11. E. H. Berger and J. E. Kerivan. Influence of the physiological noise and the occlusion effect on the measurement of the real-ear attenuation at threshold. *J. of the Acoustical Society of America*, 74(1):81–94, 1983.
12. S. Bilbao. Personal communication, 2001.
13. S. Bilbao. *Wave and Scattering Methods for the Numerical Integration of Partial Differential Equations*. PhD thesis, CCRMA, Stanford University, Stanford (CA), May 2001. available at http://ccrma-www.stanford.edu/∼bilbao/.

14. S. Bilbao and J. O. Smith. Finite difference schemes and digital waveguide networks for the wave equation: Stability, passivity and numerical dispersion. *IEEE Trans. on Speech and Audio Processing*, 2003. Accepted for publication.
15. J. Blauert. *Spatial Hearing: the Psychophysics of Human Sound Localization.* MIT Press, Cambridge, MA, 1983.
16. B. Blesser. An interdisciplinary synthesis of reverberation viewpoints. *J. of the Audio Engineering Society*, 49(10), 2001.
17. J. Borish and J. B. Angell. An efficient algorithm for measuring the impulse response using pseudorandom noise. *J. of the Audio Engineering Society*, 31(7):478–488, July 1983.
18. D. Botteldooren. Finite-difference time-domain simulation of low-frequency room acoustic problems. *J. of the Acoustical Society of America*, 98(6):3302–3308, 1995.
19. R. Bresin, A. Friberg, and S. Dahl. Toward a new model for sound control. In *Proc. Conf. on Digital Audio Effects (DAFX-01)*, pages 45–49, Limerick, Ireland, December 2001. COST-G6.
20. R. Bristow-Johnson. The Equivalence of Various Methods of Computing Biquad Coefficients for Audio Parametric Equalizers. *Audio Engineering Society Convention*, November 1994. Preprint 3906.
21. C. P. Brown and R. O. Duda. A structural model for binaural sound synthesis. *IEEE Trans. on Speech and Audio Processing*, 6(5):476–488, September 1998.
22. D. S. Brungart. Near-field virtual audio display. *Presence*, 11(1):93–106, February 2002.
23. D. S. Brungart, N. I. Durlach, and W. M. Rabinowitz. Auditory localization of nearby sources. ii. localization of a broadband source. *J. of the Acoustical Society of America*, 106:1956–1968, 1999.
24. G. Campos and D. M. Howard. A parallel 3d digital wave guide mesh model with tetrahedral topology for room acoustic simulation. In *Proc. Conf. on Digital Audio Effects (DAFX-00)*, pages 73–78, Verona, Italy, December 2000. COST-G6.
25. G. Cariolaro. *La Teoria Unificata dei Segnali.* UTET, Bologna, Italy, 1991.
26. H. C. Chan, Y. K. Cheung, and C. W. Cai. Exact solution for vibration analysis of rectangular cable networks with periodically distributed supports. *J. of Sound and Vibration*, 218(1):29–44, 1998.
27. Y. K. Cheung, H. C. Chan, and C. W. Cai. Natural vibration analysis of rectangular networks. *Int. J. of Space Structures*, 3(3):139–152, 1988.
28. Y. K. Cheung, H. C. Chan, and C. W. Cai. Dynamic response of an orthogonal cable network subjected to a moving force. *J. of Sound and Vibration*, 156(2):337–347, 1992.
29. J. M. Chowning. The simulation of moving sound sources. *J. of the Audio Engineering Society*, 19(1):2–6, 1971. Reprinted in the Computer Music Journal, June 1977.
30. C. Christopoulos. *The Transmission-Line Modelling Method.* Institute of Electrical and Electronics Engineers, 1995.
31. M. Cohen. Emerging and exotic auditory interfaces. In *Proc. Int. Conf. on Auditory Display*, pages 93–97, Kyoto, Japan, July 2002.
32. P. D. Coleman. Failure to localize the source distance of an unfamiliar sound. *J. of the Acoustical Society of America*, 34:345–346, 1962.
33. A. C. Constantinides. Spectral transformations for digital filters. *Proc. IEE*, 117:1585–1590, 1970.
34. P. R. Cook. *Real Sound Synthesis for Interactive Applications.* A. K. Peters, L.T.D., 2002.
35. J. Dattorro. Effect design – part 1: Reverberator and other filters. *J. of the Audio Engineering Society*, 45(19):660–684, September 1997.

36. F. Deboni. *Modelli per la sintesi di spazi acustici basati su reticoli waveguide filtrati*. PhD thesis, University of Verona, Dept. of Computer Science, Verona, Italy, October 2002. In Italian.
37. B. L. Dhoopar, P. C. Gupta, and B. P. Singh. Vibration analysis of orthogonal cable networks by transfer matrix method. *J. of Sound and Vibration*, 101(4):575–584, 1985.
38. Y. Ding and D. Rossum. Filter morphing of parametric equalizers and shelving filters for audio signal processing. *J. of the Audio Engineering Society*, 43(10):821–826, 1995.
39. A. Dix, J. Finlay, G. Abowd, and R. Beale. *Human-Computer Interaction*. Prentice-Hall, Englewood Cliffs, NJ, 2nd edition, 1998.
40. Dolby headphone. Web page, January 2002. Web published at www.dolby.com.
41. R. Duda and W. Martens. Range dependence of the response of a spherical head model. *J. of the Acoustical Society of America*, 104(5):3048–3058, November 1998.
42. D. E. Dudgeon and R. M. Mersereau. *Multidimensional Digital Signal Processing*. Prentice Hall, Englewood Cliffs, NJ, 1984.
43. N. I. Durlach, B. G. Shinn-Cunningham, and R. M. Held. Supernormal auditory localization. i. general background. *Presence: Teleoperators and Virtual Environments*, 2(2):89–103, 1993.
44. J. P. Ellington and H. McCallion. The free vibrations of grillages. *J. of Applied Mechanics – Trans. of the ASME*, 26:603–607, November 1959.
45. T. Engen. Psychophysics. i. discrimination and detection. ii. scaling. In J. K. Kling and L. A. Riggs, editors, *Woodworth & Schlosberg's Experimental Psychology*, pages 11–86. Methuen, London, 3rd edition, 1971.
46. F. Fabbri. *Elettronica e musica*. Fratelli Fabbri, Milan, Italy, 1984.
47. A. Fettweis. Wave digital filters: Theory and practice. *Proc. IEEE*, 74(2):270–327, February 1986.
48. N. H. Fletcher and T. D. Rossing. *The Physics of Musical Instruments*. Springer-Verlag, New York, 1991.
49. F. Fontana and M. Karjalainen. A digital bandpass/bandstop complementary equalization filter with independent tuning characteristics. *IEEE Signal Processing Letters*, 10(4):88–91, April 2003.
50. F. Fontana and D. Rocchesso. A new formulation of the 2D-waveguide mesh for percussion instruments. In *Proc. of the XI Colloquium in Musical Informatics*, pages 27–30, Bologna, Italy, November 1995. AIMI.
51. F. Fontana and D. Rocchesso. Physical modeling of membranes for percussion instruments. *Acustica*, 83(1):529–542, January 1998.
52. F. Fontana and D. Rocchesso. Online correction of dispersion error in 2d waveguide meshes. In *Proc. Int. Computer Music Conf.*, pages 78–81, Berlin, Germany, August 2000. ICMA.
53. F. Fontana and D. Rocchesso. Signal-theoretic characterization of waveguide mesh geometries for models of two-dimensional wave propagation in elastic media. *IEEE Trans. on Speech and Audio Processing*, 9(2):152–161, February 2001.
54. F. Fontana, D. Rocchesso, and E. Apollonio. Using the waveguide mesh in modelling 3d resonators. In *Proc. Conf. on Digital Audio Effects (DAFX-00)*, pages 229–232, Verona - Italy, December 2000. COST-G6.
55. F. Fontana, D. Rocchesso, and E. Apollonio. Acoustic cues from shapes between spheres and cubes. In *Proc. Int. Computer Music Conf.*, pages 278–281, La Habana, Cuba, September 2001.
56. M. B. Gardner. Distance estimation of 0° or apparent 0°-oriented speech signals in anechoic space. *J. of the Acoustical Society of America*, 46:339–349, 1969.

57. W. G. Gardner. *3-D Audio Using Loudspeakers*. Kluwer Academic Publishers, Boston (MA), 1998.
58. W. G. Gardner. Reverberation algorithms. In M. Kahrs and K. Brandenburg, editors, *Applications of Digital Signal Processing to Audio and Acoustics*, pages 85–131. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1998.
59. W. G. Gardner and K. Martin. HRTF measurements of a KEMAR dummy-head microphone. Technical Report 280, MIT Media Lab, Cambridge, MA, 1994.
60. W. G. Gardner and K. Martin. HRTF measurements of a KEMAR. *J. of the Acoustical Society of America*, 97(6):3907–3908, June 1995.
61. W. Gaver. How do we hear in the world? explorations in ecological acoustics. *Ecological Psychology*, 5(4):285–313, April 1993.
62. W. Gaver. Using and creating auditory icons. In G. Kramer, editor, *Auditory Display: Sonification, Audification, and Auditory Interfaces*, pages 417–446. Addison-Wesley, 1994.
63. G. Giudici. Personal communication, 2000.
64. S. J. Godsill and P. J. W. Rayner. *Digital Audio Restoration*. Springer-Verlag, London, UK, 1998.
65. G. H. Golub and C. F. Van Loan. *Matrix Computations*. The John Hopkins University Press, Baltimore, 1989.
66. D. W. Grantham. Spatial hearing and related phenomena. In B. C. J. Moore, editor, *Hearing*, Handbook of Perception and Cognition, chapter 9, pages 297–345. Academic Press, San Diego, CA, 1995.
67. D. Griesinger. The psychoacoustics of apparent source width, spaciousness and envelopment in performance spaces. *Acustica*, 83:721–731, 1997.
68. R. Guski. Auditory localization: effects of reflecting surfaces. *Perception*, 19:819–830, 1990.
69. A. Härmä. Implementation of frequency-warped recursive filters. *EURASIP Signal Processing*, 80(3):543–548, March 2000.
70. A. Härmä, M. Karjalainen, L. Savioja, V. Välimäki, U. K. Laine, and J. Huopaniemi. Frequency-warped signal processing for audio applications. *J. of the Audio Engineering Society*, 48(11):1011–1031, November 2000.
71. W. M. Hartmann and A. Wittenberg. On the externalization of sound images. *J. of the Acoustical Society of America*, 99(6):3678–3688, 1996.
72. C. Hendrix and W. Barfield. The sense of presence within auditory virtual environments. *Presence: Teleoperators and Virtual Environments*, 5(3):290–301, 1996.
73. P. A. Houle and J. P. Sethna. Acoustic emission from crumpling paper. *Physical Review E*, 54(1):278–283, July 1996.
74. P. Huang, S. Serafin, and J. O. Smith. Modeling high-frequency modes of complex resonators using a waveguide mesh. In *Proc. Conf. on Digital Audio Effects (DAFX-00)*, pages 269–272, Verona, Italy, December 2000. COST-G6.
75. J. Huopaniemi, L. Savioja, and M. Karjalainen. Modeling of reflections and air absorption in acoustical spaces: a digital filter design approach. In *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pages 19–22, New Paltz (NY), October 1997. IEEE.
76. Insert earphone. Web page, January 2002. Web published at www.etymotic.com.
77. J.-M. Jot. An analysis/synthesis approach to real-time artificial reverberation. In *Proc. Int. Conf. on Acoustics, Speech, and Signal Processing*, pages 249–252, San Francisco, CA, 1992. IEEE.
78. J.-M. Jot. Real-time spatial processing of sounds for music, multimedia and interactive human-computer interfaces. *Multimedia Systems*, 7:55–69, 1999.
79. J.-M. Jot and A. Chaigne. Digital delay networks for designing artificial reverberators. In *Audio Engineering Society Convention*, Paris, France, February 1991. AES.

80. T. Kailath. *Linear Systems*. Prentice-Hall, Englewood Cliffs, 1980.
81. J. F. Kaiser. Digital filters. In F. Kuo and J. F. Kaiser, editors, *System Analysis by Digital Computers*. Wiley, New York (NY), 1966.
82. G. S. Kendall. The decorrelation of audio signals and its impact on spatial imagery. *Computer Music Journal*, 19(4):71–87, 1995.
83. O. Kirkeby, P. A. Nelson, H. Hamada, and F. Orduña-Bustamante. Fast deconvolution of multichannel systems using regularization. *IEEE Trans. on Speech and Audio Processing*, 6(2):189–194, March 1998.
84. G. Kramer. *Auditory Display: Sonification, Audification, and Auditory Interfaces*. Addison-Wesley, Reading, MA, 1994.
85. A. Krokstad, S. Strøm, and S. Sørsdal. Calculating the acoustical room response by the use of a ray tracing technique. *Journal of Sound and Vibration*, 8:118–125, 1968.
86. M. Kubovy and D. Van Valkenburg. Auditory and visual objects. *Cognition*, 80:97–126, 2001.
87. A. Kulkarni and H. S. Colburn. Role of spectral detail in sound-source localization. *Nature*, 396:747–749, Dec. 1998.
88. S. K. Kumar, S. H. Han, and K. Bowyer. On recovering hyperquadrics from range data. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 17(11):1079–1083, November 1995.
89. A. J. Kunkler-Peck and M. T. Turvey. Hearing shape. *Journal of Experimental Psychology*, 26(1):279–294, 2000.
90. H. Kuttruff. *Room Acoustics*. Elsevier Science, Essex, England, 1991. Third Ed.; First Ed. 1973.
91. H. Kuttruff. Sound field prediction in rooms. In *Proc. 15th Int. Congr. Acoustics (ICA'95)*, volume 2, pages 545–552, Trondheim, Norway, June 1995.
92. S. Lakatos, S. McAdams, and R. Caussé. The representation of auditory source characteristics: simple geometric form. *Perception and Psychophysics*, 59(8):1180–1190, 1997.
93. T. Lokki, L. Savioja, R. Väänänen, J. Huopaniemi, and T. Takala. Creating interactive virtual auditory environments. *IEEE Computer Graphics and Applications*, 22(4):49–57, 2002.
94. J. M. Loomis, R. L. Klatzky, and R. G. Golledge. Auditory distance perception in real, virtual, and mixed environments. In Y. Ohta and H. Tamura, editors, *Mixed Reality: Merging Real and Virtual Worlds*, pages 201–214. Ohmsha, Ltd., Tokio (Japan), 1999.
95. W. L. Martens. Psychophysical calibration for controlling the range of a virtual sound source: Multidimensional complexity in spatial auditory display. In *Proc. Int. Conf. on Auditory Display*, pages 197–207, Espoo - Finland, August 2001.
96. R. McGrath, T. Wildmann, and M. Fersntröm. Listening to rooms and objects. In *Proc. Conference on Spatial Sound Reproduction*, pages 512–522, Rovaniemi, Finland, 1999. AES.
97. D. H. Mershon and E. King. Intensity and reverberation as factors in the auditory perception of egocentric distance. *Perception & Psychophysics*, 18:409–415, 1975.
98. S. K. Mitra. *Digital Signal Processing. A computer-Based Approach*. McGraw-Hill, New York, 1998.
99. S. K. Mitra, K. Hirano, S. Nishimura, and K. Sugahara. Design of digital bandpass/bandstop filters with independent tuning characteristics. *Frequenz*, 44(3-4):117–121, 1990.
100. M. R. Moldover, J. B. Mehl, and M. Greenspan. Gas-filled spherical resonators: theory and experiment. *J. of the Acoustical Society of America*, 79(2):253–272, 1986.

101. J. A. Moorer. The manifold joys of conformal mapping: Applications to digital filtering in the studio. *J. of the Audio Engineering Society*, 31(11):826–840, 1983.
102. P. M. Morse and K. U. Ingard. *Theoretical Acoustics*. McGraw-Hill, New York, 1968. Reprinted in 1986, Princeton University Press, Princeton, NJ.
103. Motorola Inc., Austin, Texas. *DSP56000 24-bit Digital Signal Processor Family Manual*, 2001. Web published at http://www.motorola.com/SPS/DSP/documentation/DSP56000.html.
104. A. Mouchtaris, P. Reveliotis, and C. Kyriakakis. Inverse filter design for immersive audio rendering over loudspeakers. *IEEE Trans. on Multimedia*, 2(2):77–87, June 2000.
105. J. N. Mourjopoulos. Digital equalization of room acoustics. *J. of the Audio Engineering Society*, 42(11):884–900, November 1994.
106. D. T. Murphy and D. M. Howard. 2-d digital waveguide mesh topologies in room acoustics modelling. In *Proc. Conf. on Digital Audio Effects (DAFX-00)*, pages 211–216, Verona, Italy, December 2000. COST-G6.
107. S. T. Neely and J. B. Allen. Invertibility of a room impulse response. *J. of the Acoustical Society of America*, 66(1):165–169, July 1979.
108. S. H. Nielsen. *Distance Perception in Hearing*. PhD thesis, Aalborg University, Aalborg, Denmark, 1991.
109. J. F. O'Brien, P. R. Cook, and G. Essl. Synthesizing sound from physically based motion. In *Proc. SIGGRAPH 2001*, pages 529–536, Los Angeles, CA, August 2001.
110. A. V. Oppenheim and R. W. Schafer. *Discrete-Time Signal Processing*. Prentice-Hall, Inc., Englewood Cliffs, NJ, 1989.
111. A. Papoulis. *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, New York, 2nd edition, 1984.
112. F. Pedersini, A. Sarti, and S. Tubaro. Object-based sound synthesis for virtual environments. *IEEE Signal Processing Magazine*, 17(6):37–51, November 2000. Special Issue on Joint Audio-Video Processing.
113. C. A. Poldy. Headphones. In J. Borwick, editor, *Loudspeaker and Headphone Handbook*, pages 493–577. Focal Press, Oxford (UK), 1998.
114. M. Puckette. Pure data. In *Proc. Int. Computer Music Conf.*, pages 224–227, Thessaloniki, Greece, September 1997. ICMA.
115. M. Puckette. New public-domain realizations of standard pieces for instruments and live electronics. In *Proc. Int. Computer Music Conf.*, La Habana, Cuba, September 2001. In press.
116. A. Quarteroni. *Modellistica Numerica per Problemi Differenziali*. Springer-Verlag, Milan, Italy, 2000.
117. M. Rath. Personal communication, 2001.
118. M. Rath, D. Rocchesso, and F. Avanzini. Physically based real-time modeling of contact sounds. In *Proc. Int. Computer Music Conf.*, Göteborg, September 2002.
119. P. A. Regalia and S. K. Mitra. Tunable digital frequency response equalization filters. *IEEE Trans. on Acoustics, Speech and Signal Processing*, ASSP-35(1):118–120, Jan. 1987.
120. P. A. Regalia, S. K. Mitra, and P. P. Vaidyanathan. The digital all-pass filter: A versatile signal processing building block. *Proc. IEEE*, 76(1):19–37, January 1988.
121. D. D. Rife and J. Vanderkooy. Transfer-function measurements using maximum-length sequences. *J. of the Audio Engineering Society*, 37(6):419–444, June 1989.
122. D. Rocchesso. The ball within the box: a sound-processing metaphor. *Computer Music Journal*, 19(4):47–57, Winter 1995.
123. D. Rocchesso. Acoustic cues for 3-d shape information. In *Proc. Int. Conf. on Auditory Display*, pages 175–180, Espoo - Finland, August 2001.

124. D. Rocchesso. Spatial effects. In U. Zölzer, editor, *DAFX: Digital Audio Effects*, pages 137–200. John Wiley and Sons, Inc., New York, 2002.
125. D. Rocchesso, R. Bresin, and M. Fërnstrom. Sounding objects. *IEEE Multimedia*, 2003. in press.
126. D. Rocchesso and P. Dutilleux. Generalization of a 3d resonator model for the simulation of spherical enclosures. *Applied Signal Processing*, 2001(1):15–26, 2001.
127. D. Rocchesso and L. Ottaviani. Acoustic cues for 3-d shape information. In *Proc. of the 2001 Int. Conf. on Auditory Display*, pages 175–180, July 2001.
128. D. Rocchesso and L. Ottaviani. Can one hear the volume of a shape? In *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pages 115–118, NewPaltz, NY, October 2001. IEEE.
129. D. Rocchesso and F. Scalcon. Bandwidth of perceived inharmonicity for physical modeling of dispersive strings. *IEEE Trans. on Speech and Audio Processing*, 7(5):597–601, 1999.
130. D. Rocchesso and A. Vidolin. Sintesi del movimento e dello spazio nella musica elettroacustica. In *Atti del convegno La Terra Fertile*, L'Aquila, Italy, October 1996.
131. W. Rudmose. The case of the missing 6 db. *J. of the Acoustical Society of America*, 71(3):650–659, 1982.
132. L. Russolo. L'arte dei rumori. Pamphlet, March 1913. Web published in English at http://www.futurism.org.uk/manifestos/manifesto09.htm.
133. A. Sarti and G. De Poli. Toward nonlinear wave digital filters. *IEEE Trans. on Speech and Audio Processing*, 47(6):1654–1668, June 1999.
134. L. Savioja, J. Backman, A. Järvinen, and T. Takala. Waveguide mesh method for low-frequency simulation of room acoustics. In *Proc. 15th Int. Conf. on Acoustics (ICA-95)*, pages 637–640, June 1995.
135. L. Savioja, T. Rinne, and T. Takala. Simulation of room acoustics with a 3-D finite difference mesh. In *Proc. Int. Computer Music Conf.*, pages 463–466, Århus, Denmark, September 1994.
136. L. Savioja and V. Välimäki. Improved discrete-time modeling of multi-dimensional wave propagation using the interpolated digital waveguide mesh. In *Proc. Int. Conf. on Acoustics, Speech, and Signal Processing*, pages 459–462, Munich (Germany), April 1997.
137. L. Savioja and V. Välimäki. Reduction of the dispersion error in the triangular digital waveguide mesh using frequency warping. *IEEE Signal Processing Letters*, 6(3):58–60, March 1999.
138. L. Savioja and V. Välimäki. Reducing the dispersion error in the digital waveguide mesh using interpolation and frequency-warping techniques. *IEEE Trans. on Speech and Audio Processing*, 8(2):184–194, March 2000.
139. M. R. Schroeder. *Fractal, Chaos, Power Laws: Minutes from an Infinite Paradise*. W.H. Freeman & Company, New York, NY, 1991.
140. M. R. Schroeder and B. F. Logan. "colorless" artificial reverberation. *J. of the Audio Engineering Society*, 9:192–197, July 1961. reprinted in the IRE Trans. on Audio.
141. C. L. Searle, L. D. Braida, M. F. Davis, and H. S. Colburn. Model for auditory localization. *J. of the Acoustical Society of America*, 60(5):1164–1175, 1976.
142. Electrostatic headphones. Web page, January 2002. Web published at www.sennheiser.com.
143. J. P. Sethna, K. A. Dahmen, and C. R. Myers. Crackling noise. *Nature*, (410):242–250, March 2001.
144. B. R. Shelton and C. L. Searle. The influence of vision on the absolute identification of sound-source position. *Perception & Psychophysics*, 28:589–596, 1980.

145. B. Shinn-Cunningham, S. Santarelli, and N. Kopco. Tori of confusion: Binaural localization cues for sources within reach of a listener. *J. of the Acoustical Society of America*, 107(3):1627–1636, March 2000.
146. B. P. Singh and B. L. Dhoopar. Membrane analogy for anisotropic cable networks. *J. of Structural Division American Society of Civil Engineers*, (100):1053–1066, May 1974.
147. J. O. Smith. Music applications of digital waveguides. Technical Report STAN-M-39, CCRMA - Stanford University, Stanford, CA, 1987.
148. J. O. Smith. Physical modeling using digital waveguides. *Computer Music Journal*, 16(4):74–91, Winter 1992.
149. J. O. Smith. Principles of digital waveguide models of musical instruments. In M. Kahrs and K. Brandenburg, editors, *Applications of Digital Signal Processing to Audio and Acoustics*, pages 417–466. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1998.
150. J. O. Smith and D. Rocchesso. Aspects of digital waveguide networks for acoustic modeling applications. web published at http://www-ccrma.stanford.edu/∼jos/wgj/, December 1997.
151. S. S. Stevens. The direct estimation of sensory magnitude-loudness. *American Journal of Psychology*, 69:1–25, 1956.
152. J. Strikwerda. *Finite Difference Schemes and Partial Differential Equations*. Wadsworth & Brooks, Pacific Grove, CA, 1989.
153. J. Szczupak and S. K. Mitra. Detection, location, and removal of delay-free loops in digital filter configurations. *IEEE Trans. on Acoustics, Speech and Signal Processing*, ASSP-23(6):558–562, 1975.
154. G. Theile. Loudness: independent of the direction of the incident sound? In *Proc. of the 108th Meeting of the Acoustic Society of America*, 1984. (Abstract) Journal of the Acoustic Society of America, Suppl. 1, 76, S92.
155. N. Tsingos, T. Funkhouser, A. Ngan, and I. Carlbom. Modeling acoustics in virtual environments using the uniform theory of diffraction. In *Computer Graphics (SIGGRAPH 2001)*, Los Angeles, CA, August 2001.
156. P. P. Vaidyanathan. *Multirate Systems and Filter Banks*. Prentice Hall, Englewood Cliffs, NY, 1993.
157. S. A. Van Duyne and J. O. Smith. The 2-D digital waveguide mesh. In *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, Mohonk, NY, 1993. IEEE.
158. S. A. Van Duyne and J. O. Smith. Physical modeling with the 2-D digital waveguide mesh. In *Proc. Int. Computer Music Conf.*, pages 40–47, Tokyo, Japan, 1993. ICMA.
159. S. A. Van Duyne and J. O. Smith. A simplified approach to modeling dispersion caused by stiffness in strings and plates. In *Proc. Int. Computer Music Conf.*, pages 407–410, Aarhus, Denmark, September 1994. ICMA.
160. S. A. Van Duyne and J. O. Smith. The tetrahedral digital waveguide mesh. In *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, page 9a.6, Mohonk, NY, October 1995. IEEE.
161. S. A. Van Duyne and J. O. Smith. The tetrahedral digital waveguide mesh with musical applications. In *Proc. Int. Computer Music Conf.*, pages 19–24, Hong Kong, August 1996.
162. G. von Békésy. The moon illusion and similar auditory phenomena. *American Journal of Psychology*, 62:540–552, 1949.
163. M. Vorländer. Acoustic load on the ear caused by headphones. *J. of the Acoustical Society of America*, 107(4):2082–2088, Apr. 2000.
164. L. R. Wanger, J. A. Ferwerda, and D. P. Greenberg. Perceiving spatial relationships in computer-generated images. *IEEE Computer Graphics & Applications*, pages 44–58, May 1992.

165. 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.
166. O. Warusfel, E. Kahle, and J. P. Jullien. Relationships between objective measurements and perceptual interpretation: The need for considering spatial emission of sound sources. *J. of the Acoustical Society of America*, 93(4):2281–2282, April 1993.
167. P. Zahorik. Assessing auditory distance perception using virtual acoustics. *J. of the Acoustical Society of America*, 111(4):1832–1846, April 2002.
168. P. Zahorik. Auditory display of sound source distance. In *Proc. Int. Conf. on Auditory Display*, Kyoto, Japan, July 2002.
169. P. Zahorik. Direct-to-reverberant energy ratio sensitivity. *J. of the Acoustical Society of America*, 112(5):2110–2117, November 2002.
170. U. Zölzer and T. Boltze. Parametric digital filter structures. *Audio Engineering Society Convention*, October 1995. Preprint 3898.
171. E. Zwicker and H. Fastl. *Psychoacoustics: Facts and Models*. Springer Verlag, Berlin, Germany, 1990.

Acknowledgments

It is impossible for me to acknowledge all the researchers I met during these years, for whom I feel gratitude. Some of them are well-known personalities in their field, some others are less present in the official boards. Independently of their position and visibility in the community I found them to be as more engaging, as more sincerely involved in their research topic.

There is no reason to acknowledge here my family and relatives: they are *always* part of my life, and their presence in it is well beyond whatever role they may have played during the working out of this thesis.

My first acknowledgments go to my early teachers and colleagues in Padua: Giovanni De Poli, Roberto Bresin, Federico Avanzini, Nicola Orio, Carlo Drioli. And Gianpaolo Borin, whose level of expertise in digital audio is far beyond the little number of precious publications and talks he gifted to our community.

I cannot forget the professional and human experience I had during my research visit at the Helsinki University of Technology. At the Laboratory of Acoustics and Audio Signal Processing, for the first time in my life I was in a community where several researchers coming from different countries, cultures and backgrounds, having different positions in the academia or the industry, working in multidisciplinary scientific fields with different levels of expertise, joined together to create a unique synergy. Matti, Mark, Vesa: I had exciting meetings and talks with you, and I learned so much from your scientific, artistic and sometimes intuitive vision of sound. Cumhur, Hanna, Henri, Jyri, Lauri, Riitta, Ville: I would have just spent more time with you. I deserve a special place in my mind for all the Aku guys and, more in general, for Finland and its exotic atmosphere.

I cannot mention all my colleagues in Verona. Vittorio Murino, Andrea Fusiello, Manuele Bicego, Laura Ottaviani, just to cite those with whom I was mostly in touch. I had occasions to go beyond the scientific discussion with Linmi, Matthias and Umberto, whose guest rooms became invaluable for me more than one time. I had discussions of *real* computer science with Vincenzo Manca and Roberto Giacobazzi. I would finally mention my PhD colleagues and some of the research people in Verona: Alessandro, Arnaud, Debora, Giuditta, Graziano, Isabella, Nicola, and Nicola, for whom I hope the best. And Linda: I will never catch your theorems.

Jyri Huopaniemi and Augusto Sarti have reviewed this thesis: their notes have helped me in rearranging and improving several parts of it.

I supervised the graduation thesis of several students, both in Padua and in Verona. I hope they enjoyed working with me in the subjects I proposed. Any one of them contributed to make a topic more clear or more difficult to understand: Luca, Enzo, Federico, Fabio, Giulio.

Monique Harris Muz translated in English the citations I added in the first part of the thesis, at the beginning of each chapter. Zane Mackin gave me hints for improving the English of the Preface and the Introduction and translated the citation at the beginning of Chapter 5, whose source in English I was unable to recover.

And, finally, I cannot forget the people I met during the international activities of our group. I used to interact with them in an exciting and often funny way: Nicola Bernardini, Mikael Fernström, Marco Trevisani, Giovanni Bruno Vicario. And Bruno, Massimo, Sophia, Eoin, Kjetil, Mark. And Balazs: you will always find a hot meal in Italy.

My last, and strongest thanks go to Davide Rocchesso. We start collaborating together when he was a PhD candidate and I was a puzzled student of engineering, and an enthusiastic drummer. That time I started dealing with the audio applications of computer science. I hope not to stop working on them in the years to come.

Contestualizzazione della ricerca e riassunto dei contenuti

In questo lavoro sono state affrontate alcune questioni inserite nel tema più generale della rappresentazione di scene e ambienti virtuali in contesti d'interazione uomo-macchina, nei quali la modalità acustica costituisca parte integrante o prevalente dell'informazione complessiva trasmessa dalla macchina all'utilizzatore attraverso un'interfaccia personale multimodale oppure monomodale acustica.

Più precisamente è stato preso in esame il problema di come presentare il messaggio audio, in modo tale che lo stesso messaggio fornisca all'utilizzatore un'informazione quanto più precisa e utilizzabile relativamente al contesto rappresentato. Il fine di tutto ciò è riuscire a integrare all'interno di uno scenario virtuale almeno parte dell'informazione acustica che lo stesso utilizzatore, in un contesto stavolta reale, normalmente utilizza per trarre esperienza dal mondo circostante nel suo complesso. Ciò è importante soprattutto quando il *focus* dell'attenzione, che tipicamente impegna il canale visivo quasi completamente, è volto a un compito specifico.

A tutt'oggi lo stato dell'arte nella rappresentazione di scene acustiche virtuali caratterizzate da specifici attributi acusto-spaziali, e contenenti oggetti in forma di sorgenti di suono in grado di interagire tra loro formando eventi acustici riconoscibili, non prevede un tipo d'interazione tra sistema e utilizzatore di tipo *diretto*, in grado cioè di soddisfare a dei requisiti che possono essere riassunti nei due punti seguenti:

- la presenza di oggetti e/o eventi nella scena la cui interpretazione non richieda l'utilizzo di processi di tipo cognitivo superiore, se non addirittura un addestramento di tipo preliminare finalizzato alla comprensione del significato della scena stessa;
- la possibilità di controllare in modo intuitivo e immediato gli oggetti e gli eventi nella scena, nonchè i parametri acusto-spaziali caratterizzanti la scena nel suo complesso.

Da qualche tempo è in corso presso alcune comunità uno sforzo di ricerca per definire metodi per la caratterizzazione acustica di oggetti ed eventi nella scena, attraverso una loro realistica integrazione nel contesto. Queste ricerche fra l'altro hanno portato a modelli in grado di rappresentare fenomeni fisici corrispondenti a eventi salienti della scena stessa, quali ad esempio il rotolamento e lo schiac-

ciamento di oggetti su una superficie, l'impatto e lo sfregamento fra oggetti, la rottura eccetera.

È evidente che simili modelli soddisfano i requisiti espressi ai punti sopra. Infatti, la comprensione dell'evento sonoro è in questi casi puramente *ecologica*, dunque immediata. Parimenti, il controllo degli stessi modelli è ben fondato e, di fatto, intuitivo, in quanto dettato dalla dinamica del fenomeno fisico causante l'evento.

D'altra parte esiste la questione, accennata in precedenza, della contestualizzazione di tipo spaziale degli eventi appena descritti. In tal senso occorre ricordare che un soggetto udente percepisce la posizione angolare e la distanza degli oggetti relative alla posizione d'ascolto.

Contrariamente alla modellizzazione dei suoni emessi delle sorgenti acustiche, oggetto di ricerche riconducibili alla sintesi di suoni per l'interazione uomo-macchina, viceversa la *spazializzazione* delle stesse sorgenti, ovvero la loro collocazione nello spazio relativamente al punto d'ascolto, è oggetto di studi di acustica e di psicoacustica volti a comprendere le motivazioni psicofisiche della nostra percezione acustica dello spazio. Di conseguenza, poco è stato fatto fino a questo momento per individuare dei modelli per la sintesi di attributi spaziali, in grado di aggiungere all'informazione sulla natura degli oggetti della nuova informazione, *ancora oggettiva*, volta alla specificazione della loro posizione nello spazio.

Un approccio basato sulla rappresentazione degli attributi acusto-spaziali dell'oggetto sonoro trova un contesto applicativo qualificante in particolare nella riproduzione della distanza tra oggetto e ascoltatore: rispetto a questo problema, infatti, è stato dimostrato che la componente soggettiva dovuta alla morfologia dell'ascoltatore riveste un ruolo percettivo secondario a confronto delle caratteristiche oggettive dello scenario. In altre parole, la riproduzione della distanza può essere inglobata all'interno della rappresentazione dell'oggetto e, come tale, un modello volto a rappresentare acusticamente la distanza può essere reso soggetto ai due requisiti inizialmente imposti.

Intendiamo dunque mettere a punto un modello per riprodurre acusticamente la distanza degli oggetti sonori in una scena, il quale sia in grado di rappresentare l'informazione sulla distanza in modo immediatamente fruibile per l'utilizzatore. Nel contempo desideriamo che lo stesso modello sia controllabile in parametri di riscontro immediato.

È chiaro che il soddisfacimento di questi requisiti si traduce, a livello di prestazioni dell'interfaccia, nella presenza di nuovi elementi informativi presenti nel canale audio, i quali determinano un aumento assoluto della banda d'interazione disponibile e, più in generale, un aumento del coinvolgimento dell'utilizzatore e del realismo dell'ambiente virtuale.

Il modello è stato realizzato supponendo di disporre di un tradizionale sistema per personal computer per la presentazione dell'audio, in normali condizioni di utilizzo dell'interfaccia, e tenendo conto del modo in cui un soggetto normalmente udente è in grado di percepire acusticamente la distanza relativa dall'oggetto sonoro. I risultati ottenuti negli esperimenti attestano le promettenti prestazioni del modello sviluppato.

L'attesa disponibiltà di parametri di controllo oggettivi ha richiesto di realizzare il modello adottando uno schema numerico particolarmente adatto alla simu-

lazione di domini di propagazione d'onda distribuiti nello spazio tridimensionale, chiamato *Waveguide Mesh*. L'adattamento dello schema alle condizioni imposte dal modello si è tradotto in una serie di modifiche *ad hoc* e di migliorie apportate allo schema stesso, nell'ottica di una sua ottimizzazione alla rappresentazione dell'ambiente virtuale proposto.

In parallelo, sono state studiate anche alcune problematiche di produzione e di presentazione dell'audio nell'interfaccia. Per quanto riguarda la produzione si è addivenuti a un modello per la sintesi di eventi di *fratturazione*, utile alla generazione di suoni ecologici quali il calpestio e lo schiacciamento. Per quanto riguarda la presentazione è stato messo a punto un sistema innovativo per la realizzazione di funzioni di equalizzazione inversa, utile nella rimozione di artefatti presenti nel segnale audio dovuti a distorsioni causate dal sistema di presentazione del messaggio acustico e dall'ambiente d'ascolto in cui lo stesso messaggio viene presentato.

Più precisamente, gli elementi innovativi contenuti in questo lavoro sono riassumibili nei punti seguenti:

- è stata portata a termine un'analisi spazio-temporale della *Waveguide Mesh*, in conseguenza della quale è stato possibile determinare la banda utile dei segnali ottenuti come risposta dello schema numerico al variare della geometria del reticolo, anche nel caso in cui il reticolo presenti delle lacune in corrispondenza dei suoi punti nodali;
- è stata modellata una versione modificata della *Waveguide Mesh* rettangolare, la quale produce risposte all'impulso contenenti la stessa quantità di informazione offerta dalle risposte del modello originale, pur necessitando di risorse di calcolo e di memoria dimezzate;
- è stata proposta un'interpretazione fisica dell'errore numerico, cosiddetto di *dispersione*, prodotto dalla *Waveguide Mesh*. Detta interpretazione è basata su risultati ottenuti nello studio di strutture, note come *cable networks*, che costituiscono la controparte meccanica del modello;
- è stata fornita una parametrizzazione di modelli di filtri, noti come *Digital Waveguide Filter*, utili a modellare il bordo della *Waveguide Mesh* nel caso in cui si stiano simulando superfici di riflessione d'onda acustica reali;
- è stata studiata una formulazione di tipo "scattered" del bordo della *Waveguide Mesh* utile alla modellizzazione delle riflessioni multiple d'onda, quali quelle che avvengono normalmente nelle superfici reali;
- è stata modificata la struttura della *Waveguide Mesh* triangolare in modo tale da minimizzarne l'errore di dispersione. Per inciso ciò ha richiesto di risolvere da un punto di vista generale il problema della propagazione di un segnale all'interno di una rete lineare di filtri contenente cicli senza ritardo, notoriamente non calcolabili utilizzando le classiche procedure di elaborazione del segnale. Detta soluzione si traduce in una procedura che fornisce automaticamente il sistema implicito di equazioni derivato dalla rete stessa;
- è stata definita una classe di geometrie per una cavità tridimensionale risonante, in modo da studiare gli aspetti percettivamente interessanti contenuti negli spettri delle risposte ottenute dalla stessa cavità man mano che la sua geometria varia, lungo la classe, dalla forma sferica alla forma cubica. Ciò per indagare la percezione della forma degli oggetti a partire da indizi di tipo sonoro;

- è stato messo a punto un modello per la sintesi di indizi di *profondità acustica*, al fine di evocare in un soggetto il senso della distanza a partire da informazioni audio. È questo il punto più qualificante della ricerca svolta, nel quale sono stati compendiati i punti precedenti;
- è stata portata a termine una campagna di esperimenti volta a valutare il modello per la resa della distanza in diverse condizioni d'ascolto, facendo uso ora di cuffie audio, ora di altoparlanti. Da questi esperimenti sono emersi risultati incoraggianti, tali da lasciar intravedere un effettivo utilizzo del modello in interfacce uomo-macchina e in applicazioni di realtà virtuale di tipo convenzionale e avanzato;
- è stata sviluppata una struttura innovativa per una classe di filtri digitali nota come *equalizzatori parametrici*, la cui specificità consiste nella semplicità d'inversione della loro funzione di trasferimento. Questa proprietà torna utile nella compensazione di echi e risonanze indesiderate contenute nel messaggio audio;
- desumendo la fisica della fratturazione da ricerche precedenti, si è addivenuti a un metodo di sintesi di suoni *particellari*, la cui sovrapposizione secondo ben definite regole statistiche ha permesso di sintetizzare suoni di calpestio e di schiacciamento. Il modello di sintesi ottenuto risulta controllabile in parametri di riscontro immediato quali le dimensioni dell'oggetto schiacciato, la forza di schiacciamento e la resistenza del materiale schiacciato. Attraverso l'uso di tale metodo si sono potuti definire dei suoni campione, successivamente applicati al modello per la resa della distanza per la creazione di esempi audio opportuni.

La tesi, in lingua inglese, è stata organizzata in due parti. La prima parte contiene un'introduzione e un compendio generale del lavoro complessivo svolto. La seconda parte presenta in ordine cronologico alcuni lavori scritti dall'autore, pubblicati o sottoposti a revisione.

Attraverso questa organizzazione si è cercato di rendere più agevole l'accesso al contenuto complessivo anche al lettore meno interessato a entrare nello specifico dei singoli argomenti. Viceversa, il lettore desideroso di approfondire le tematiche proposte è invitato a considerare una lettura della seconda parte.

Gli esempi audio sono proposti nel sito web dell'autore.