Diploma Thesis

A Tangible User Interface for Playing Virtual Acoustics

Author:
Birgit Gasteiger

Institute of Electronic Music and Acoustics (IEM)
University of Music and Performing Arts Graz

Assessor:
Univ.Prof. Dr.phil. Gerhard Eckel

Supervisor:
Ao.Univ.Prof. DI Winfried Ritsch

Graz, March 17, 2010

Acknowledgments

I like to thank my advisor Winfried Ritsch for his support and for his great idea to work on this theme. I wish to thank Gerhard Eckel for his expertise and for the comentoring of this thesis. I also wish to thank Alois Sontacchi for providing scientific hints and valuable suggestions.

Special mention goes to Markus Reichhartinger, who encouraged me in the choice of an engineering study. Furthermore, I wish to thank Gerda Strobl, Carola Ries, Christina Leitner, Martin Rohrmoser, Georg Holzmann, Friedrich Schäfer and Markus Guldenschuh, for accompanying me during my years of study. Many thanks also to Brigitte Bergner.

Especially I like to thank my family for giving me their lifelong support in all my intentions and for their motivation in gaining knowledge.

I like to dedicate this work to my partner Benjamin. I wish to thank him for his marvelous support in every sense, which is beyond words, and for being there for me.

Abstract

In this thesis the development of a digital musical instrument (DMI) playing virtual architectural spaces in the manner of a musical instrument is presented. The structure of the instrument is based on the implementation of an Auditory Virtual Environment (AVE) approach. The AVE approach comprises a geometric room model enabling real-time acoustic modeling of rectangular room geometries and supporting a dynamic modification of boundaries and sources. The computation of early reflection paths is derived by the implementation of an image-source model. A statistical model provides the simulation of reverberation and is based on a feedback delay network (FDN). Higher Order Ambisonics (HOA) is employed for the reproduction of the sound field. In order to enable instrumental interaction a tabletop tangible interface (TUI) providing tangible and direct-touch interaction is built up. The results of an objective measurement evaluating the auditory quality of the environment are presented.

Zusammenfassung

Diese Arbeit beschreibt die Entwicklung eines digitalen Musikinstrumentes auf der Basis einer tangiblen, tischbasierten Benutzerschnittstelle (engl. Tabletop Tangible User Interface), welches es ermöglicht künstlich erzeugte Raumakustik in musikalischer Weise zu steuern. Die Grundstruktur des Instrumentes basiert auf der Erzeugung einer virtuellen Hörumgebung (engl. Auditory Virtual Environment) mittels geometrischem Raumakustikmodell. Das Modell unterstützt eine Raumakustiksimulation in Echtzeit, für rechteckige Geometrien, veränderliche Grenzflächen und Schallquellenpositionen. Die Berechnung der frühen Reflexionen wird mittels Spiegelquellenmodell realisiert. Die Simulation des Nachhalles erfolgt durch die Implementation eines Rückkopplungsnetzwerkes (engl. Feedback Delay Network). Die Reproduktion des Schallfeldes in der künstlich geschaffenen Hörumgebung erfolgt mittels Higher Order Ambisonics. Um die subjektive Empfindung in der Hörumgebung zu schätzen, werden unterschiedliche Raumeinstellungen simuliert und mittels objektiver Messungen evaluiert.

Contents

Part I
Space as musical instrument

1 Introduction

The motivation of using space as a musical instrument derives from the perceivable influence of room acoustics on the enclosed sound source. Ando [And09] terms the concert hall as second musical instrument contributing to the performance of a musical instrument on stage. He emphasizes the spatial and temporal aspects of the sound field and their contribution to the musical expression. These perceptual aspects are considered at the composition of a piece and are exploited as compositional tool in traditional music and in computer music.

The motivation in this thesis is to simulate virtual room acoustics and to derive a setup which enables playing a virtual space like a musical instrument. For the implementation of the instrument an auditory virtual environment (AVE) approach, introduced by Blauert [Bla05] is proposed to provide the basic structure of the instrument. An AVE approach concentrates on the simulation of the auditory part of a virtual environment. Thus a room acoustic model defines the virtual environment and with the placement of sound sources the acoustic properties of the environment get audible.

In order to provide a control possibility enabling musical interaction with the virtual environment, in the course of this thesis a tabletop tangible user interface TUI is implemented. The concept of tangible interaction [Ull95] in the combination of direct-touch interaction, which provides the control of digital information by the use of tangible objects on the surface of the interface, is employed. The development of the digital functionality of the tangible objects in order to provide a dynamic modification of the virtual architecture is an additional contribution of this thesis. Furthermore a visual feedback, which includes a scene graph for the representation of the virtual environment and which provides dynamic visual information about the instrument state is developed.

In the fist part of the thesis, in chapter 2 the musical background of the composition including the spatial character of the performance space is presented. In chapter 3 the definition of an AVE approach and the basic structure of the instrument are introduced.

The second part provides the theoretical background which provides the basement for the development of the instrument. Chapter 4, 5, 6 and 7 provide the theoretical basement of room acoustics, human auditory perception in spaces and the objective evaluation of the subjective perception, for the development of the perceptual quality of the AVE. In chapter 8 computational efficient room acoustic modeling methods are discussed. Chapter 9 provides an overview of spatial sound reproduction techniques. The concept of tangible interaction and the setup of a tabletop tangible interface is described in chapter 10 and 11.

In the third part the practical contributions, including the development of the instrument functionality, software and hardware implementation and measurement results are presented in chapter 12, 13, 14, 15 and 16.

2 Paradigms of composition with space

In the following a music style and several compositions using the concert hall or performance space as second instrument are listed below:

The **Venetian polychoral style** [Wik09] was a type of music which occurs in the late 16th century in Venice. In this style choirs were alternately singing in spatially distributed formation in the Basilica San Marco di Venezia. It evolves due to the architectural character of the basilica which contributes to a peculiar sound perception. The composers, aware of the delay caused by the distance between the oppositing choirs used this phenomenon as special effect.

Poème électronique [Tre96] [ZLM+05] is a electronic music piece of Edgar Varèse, written for the Philips Pavilion at the Brussels World's Fair center in 1958. In the piece a complex sound spatialization scheme, synchronized with film projections, reproduced through a sound projection based on the style of the Acousmonium, which is sound diffusion system designed in 1974 by Francois Bayle. The positioning of hundreds of speakers enabled the exploitation of the unique physical layout of the pavilion and Varèse e.g. made use of sending the sound up and down the walls.

I am sitting in a room [Luc05] is a musical piece of Alvin Lucier composed in 1970. Lucier investigates in his artistic work the physical nature of sound and observes the effect of sound on the listener. He demonstrates the sonic character of physical phenomenons and aims in particular at the visualization of sound. With this particular piece Lucier demonstrates the impact of a room on the enclosed sound source. For this purpose Lucier sits in a small room and records himself reading a text passage. He reproduces the recording via speakers and records it again many times until the sound of his voice is transformed due to the acoustic properties of the room.

3 Auditory Virtual Environment (AVE)

The creation of an auditory virtual environment (AVE) approach, [Bla05, chap. 11], provides the basic structure of the instrument. A virtual environment aims at creating situations in which humans have perceptions not corresponding to the physical environment but to a virtual one. In this environment the most important perceptions are vision, audition, proprioception and tactility. An auditory virtual environment focuses on the auditory component of the environment which is regarded independently from the other components. An AVE system comprises sound source, environment and a listener model which can be implemented using three different approaches:

- The **authentic approach** aims at the reproduction of an existing real environment and at evoking an auditory perception indistinguishable from the perception in a corresponding real environment.
- The **plausible approach** aims at creating an auditory perception of a virtual environment which could have occurred in a real environment. In this application only those features which are relevant for a particular application need to be considered.
- The **creational approach** aims at creating auditory events for which no restrictions of authenticity or plausibility are assumed.

3.1 AVE modules

The main modules of an AVE include sound source, environment, reproduction format and signal processing. The block diagram of the four main modules and data flow paths of a typical auditory-virtual-environment system is illustrated in figure 1.

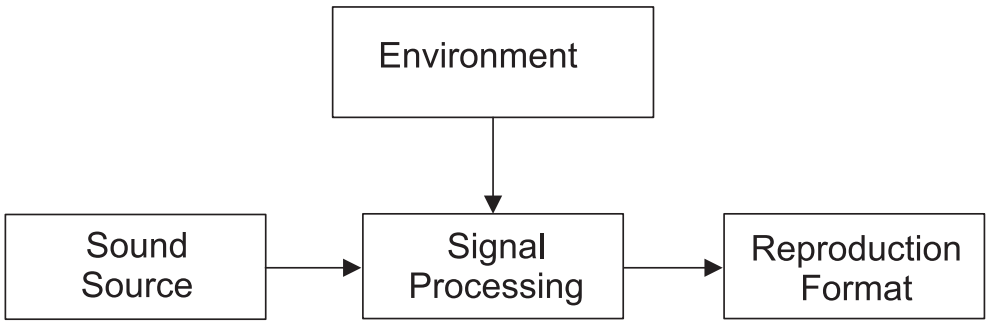


Figure 1: Modules of a auditory virtual environment system [Bla05, p.279]

The **sound source** signals in an AVE can be described in terms of generation method, time domain characteristics, spectral characteristics, directivity and whether the source is moving or in a static position. The aspects are discussed in chapter 12.4.

In the **environment model** the physical properties of the environment and their impact on the perception of the sound source, e.g. perception of reflection, are modeled. Physical room acoustic modeling methods are divided into wave-based, geometric-based and statistics-based methods. With regard to the implementation of

an AVE in musical context these methods are further discussed in terms of real-time implementation in chapter 8.

The task of the **signal processing** is to process the sound-source signals based on the physical room model and to output the result in a specified reproduction format. In a geometrical-acoustics approximation, see chapter 4.2, the room acoustic model calculates the propagation path from the sound source to the receiver. For each propagation path three processing steps are achieved: (a) delay corresponding to length of propagation path, (b) spectral modification due to one or multiple reflections and directional source characteristics and (c) the direction of incidence according to the listener's head. The signal processing for the instrumental approach is described in chapter 8.

Reproduction formats in [Bla05, chap. 11.2.4], are divided into formats using head related transfer functions (HRTFs), which are described in chapter 5.1, or do not use HRTFs. Approaches using HRTFs further distinguish headphone-based systems and loudspeaker-based systems. Reproduction formats not employing HRTFs are "vector-based amplitude panning" VBAP, Ambisonics and "wave-field synthesis". These reproduction formats differ whether the auditory events are located over a two-dimensional or three-dimensional plane and whether the physical reproduction of the sound field is derived. In chapter 9 sound spatialization methods are further discussed in terms of affording a high quality in implementation of the auditory virtual environment.

3.2 Quality assessment of auditory virtual environments

The quality measures of virtual environments are distinct from the quality measures of a real environment and are of importance in the design phase of an AVE. Based on the knowledge about human auditory perception a number of simplifications, with regard to save processing cost and memory requirements but without a degrading the perceived quality, can be undertaken. Pellegrini [Pel01] describes meaningful quality measures to assess the perceived quality for the user within an AVE implementation. In the following quality measures, important for the implementation of the instrument are described more detailed.

3.2.1 Authentic vs. plausible approach

With the precondition of plausibility of an AVE approach, a set of demanded quality features for a given specific application is defined. The application specific, relevant quality features can be determined in auditory tests and can thus be independently observed. Furthermore the reproduction of the AVE can be simplified since the purpose of the application is known and thus only relevant quality features need to be reproduced. This leads to the reduction of processing cost and memory requirements. The relevant quality features depend on the purpose of the application and are in the context of an instrumental purpose defined at the end of this chapter in 3.3.

3.2.2 Presence

In most applications subjects should experience presence in order to fulfill a specific task intuitively. Presence is the sense of being inside the virtual environment. The

sense of presence is provided by the possibility of interaction between subject and environment. Furthermore the subjective presence is supported by the physical presence provided by the physical presence of the listener in the same physical space in which the auditory event occurs.

3.2.3 Usability

Usability is described as a combination of effectiveness, efficiency, and acceptance. Effectiveness rates the accuracy and completeness with which specified users can achieve specified goals in particular environments. Efficiency describes the resources expended in relation to the accuracy and completeness of goals achieved. It is measured in general by the mean time taken to achieve the task. The acceptance describes the subjective level of satisfaction of the user in using the product.

3.2.4 Dynamic aspects of quality in AVEs

A dynamic auditory virtual environment provides a change of the properties of the source, environment or listener over time. Changes may be induced from outside or due to interaction of user and environment. In contrast to a real environment a delay between action and corresponding reaction exists and the maximal allowable delay for the reaction is defined by the application. This aspect is discussed in chapter 10.2 in the context of a musical interaction. From a technical point-of-view the latency, the update rate and the resolution of dynamic changes are influencing the dynamic quality of the system. Thus the delays introduced by the virtual-environment system should be kept below the just-noticeable-differences for human perception or, if not achievable, below an annoying threshold which is depending on the task of the approach.

3.3 Conclusions concerning the implementation of the AVE

For the purpose of the implementation of a musical instrument, the realization of a plausible AVE approach is proposed. The features of the AVE are defined in order to provide an adequate interaction the environment and to derive an appealing and dynamic auditory output of the instrument. Thus the desired physical features are specified as follows:

- A dynamic AVE is supported by the application of static and dynamic sound sources, room geometries and room acoustic properties of the environment.
- The maximal allowable delay in the reaction of the AVE system is limited conform to the requirements of playing a musical instrument.
- The subjective presence of the listener inside the auditory virtual environment should be supported by the presence of the listener in the physical space where the auditory events occur.

Part II
Theoretical background

4 Room acoustics

Room acoustics is an important branch of acoustics and deals with the description of sound fields in closed spaces. Room acoustics is described through three different approaches: (1) wave theory based room acoustics, (2) statistical room acoustics and (3) geometrical room acoustics.

4.1 Wave-theory-based room acoustics

The wave-theory-based approach is based on the wave equation. The derivation of the wave equation and the description of a plane harmonic wave according to Kuttruff [Kut09] are discussed in the following chapters. Furthermore basic relations concering sound propagation, reflection, absorption and modes in enclosed spaces are described.

4.1.1 Wave equation

In a sound wave the particles of the medium are displaced in terms of time and space. A sound wave thus can be completely described by the velocity v of the particle displacement. At first the the momentum of the sound pressure is expressed by:

$$grad\ p = -\rho_0 \frac{1}{c} \frac{\delta v}{\delta t} \tag{1}$$

with p sound pressure, v is the particle velocity vector, t is the time and ρ_0 is the static value of gas density. Furthermore the conservation of mass leads to

$$\rho_0\ div\ v = -\frac{\delta \rho}{\delta t} \tag{2}$$

with ρ as total density including the variable part $\rho = \rho_0 + \delta\rho$. The elimination of particle velocity v and the variable part $\delta\rho$ results in the wave equation:

$$\Delta p = \frac{1}{c} \frac{\delta^2 p}{\delta t^2} \tag{3}$$

where

$$c^2 = \kappa \frac{p_0}{\rho_0} \tag{4}$$

with κ is the adiabatic or isentropic exponent (for air $\kappa = 1.4$).

4.1.2 Plane harmonic wave

In harmonic waves the time and space dependence of the acoustical quantities, such as the sound pressure, follow a sine or cosine function. The expression of a plane harmonic wave is represented by:

$$p(x,t) = \hat{p}cos[k(ct - x)] = \hat{p}cos(\omega t - kx) \tag{5}$$

with arbitrary constants of $\hat{p}$ and k. The quantity k = ω/c is the propagation constant or the wave number of the wave and $\hat{p}$ is the amplitude. The angular frequency ω = kc is related to the temporal period T = $2\pi/\omega$. The wavelength of the harmonic wave is $\lambda = 2\pi$/k. The wavelength is related to the angular frequency by λ = 2πc/ω = c/f, where f denotes the frequency in Hertz of the vibration.

4.1.3 Sound propagation

An important law of sound propagation in free space is the dependence of sound intensity and sound pressure level on the propagation distance. The intensity of a spherical point source decreases with $1/r^2$, where r denotes the distance to the sound source.

4.1.4 Reflection and absorption

If a plane wave strikes a wall one part of the sound intensity of the incident wave is reflected [Kut09]. At the point of incidence on the wall the reflected wave emanates with different phase and amplitude depending on angle of incidence and frequency. The changes in phase and amplitude are indicated by the complex reflection factor R which depends on the property of the wall:

$$R = |R|^{(i\chi)}. \tag{6}$$

The absolute value and the phase angle depend on the frequency and the direction of the incident wave.

Since the intensity of a plane wave is proportional to the square of the pressure amplitude, the reflected wave is smaller by a factor $|R|^2$ than the incident wave and the part 1-$|R|^2$ of the incident energy is lost during reflection. This energy loss is described by the absorption coefficient α:

$$\alpha = 1 - |R|^2. \tag{7}$$

If the wall is totally absorbent α derives a maximum value of one. In the case of "rigid" walls, R = 1, which means in-phase reflection, χ= 0.

4.1.5 Modes in enclosed spaces

If the geometry of a room is large compared to the wavelength the correct calculation of the sound field is based on the wave equation [Vor08]. With a harmonic volume source excitation q = $\hat{q}e^{j\omega t}$ the wave equation is transformed into the Helmholtz equation. With the boundary conditions concerning geometry and impedances the equation quotes as:

$$\Delta p + k^2 p = -jkZ_0 q \tag{8}$$

where $Z_0 = \rho_0$ c.

Discrete values of k_{xyz} 'eigenvalues' are determined such that they fulfill the differential equation and boundary conditions. The eigenfrequencies f_{xyz} are derived by:

$$f_{xyz} = \frac{c}{2\pi} k_{xyz}. \tag{9}$$

For example a rectangular room of the dimensions l_x, l_y and l_z and wall impedance Z=∞, hard wall, the eigenfrequencies are determined by the following equation:

$$f_{xyz} = \frac{c}{2}\sqrt{(\frac{x}{l_x})^2 + (\frac{y}{l_y})^2 + (\frac{z}{l_z})^2} \tag{10}$$

with x, y and z $\in N_0$. The density of the modes (eigenfrequencies) increases with the frequency independant of the room shape.

4.2 Geometrical room acoustics

In geometrical room acoustics sound waves are approximated by sound rays [Kut09]. The description of sound propagation corresponds to geometrical ray optics. The following simplifications and idealizations are assumed:

- The wavelength of sound is small compared to the dimensions of the surface and large compared to the roughness.
- The intensity of rays emitted by the sound source decreases with a factor $1/r^2$, where r=distance to point of origin.
- Diffraction phenomenons are disregarded as sound propagation in straight lines is assumed.
- Interference is not considered, in case of superposition of sound rays, phase relations are neglected, instead energy densities and intensities are added.

4.2.1 Specular reflection

If a sound ray strikes a plane and smooth wall it is reflected at the point where the arriving ray intersects the surface. The angle between the incident sound ray equals angle of the reflected ray. The vector notation of this reflection law results in:

$$u' = u - 2(un) \cdot n \tag{11}$$

where u and u' are unit vectors pointing in the direction of the incident and the reflected sound ray and n denotes the normal vector to the surface. The reflection is illustrated in figure 2.

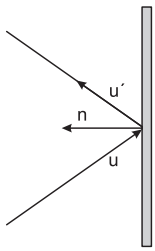


Figure 2: Angle of reflection [Kut09, p. 103]

If a point source A is reflected at a plane wall, the reflection is supposed to be emanated from an image source A' behind this wall. The virtual source A' is located

on a perpendicular line, at the same distance to the wall as the original source. Due to this position the reflection path from the original source to the listening position R can be found. As stated in chapter 4.1.4, during the reflection a part of the sound energy is not reflected and thus the image source can be supposed to emit a sound power reduced by the factor 1 - α. If the sound ray strikes a second wall the mirroring process is repeated and higher order reflections are created. The total number of images is derived by:

$$N(i) = N\frac{(N-1)^i - 1)}{N-2}. \tag{12}$$

Different traveling distances determine different delays and strengths of the reflected rays. The signal at the receiving point is derived by adding the sound intensities of all reflections. The temporal structure of the reflections arriving at the listening position is divided into direct sound, early reflections and reflections with rapidly increasing density but with a decaying energy, which are perceived as reverberation.

4.3 Statistical room acoustics

In statistical room acoustics the acoustical description of a room is based on the relation of inserted sound energy and energy absorption [Kut09]. The energy balance provides information about the decay process and the energy density of a stationary sound field, which are the central subjects in statistical room acoustics. In statistical room acoustics a diffuse sound field which assumes uniform distribution of energy density within the room, without the presence of room resonances. Relevant parameters in these considerations are the room volume V, the surface S and the degree of sound absorption α. In this context the room geometry is not considered as relevant.

4.3.1 Energy density

In order to derive a law for the description of the sound decay within a room the energy balance is calculated. A sound source is assumed to feed the acoustical power P(t) into a room. With the "'equivalent absorption area"' A_{equ} and a constant acoustical power P the steady state energy density ω is [Kut09, p. 130] derived by:

$$\omega = \frac{4P}{cA_{equ}}. \tag{13}$$

A_{equ} is calculated as follows:

$$A_{equ} = \sum_{i=1}^{n} \alpha_i \cdot S_i \tag{14}$$

where n is the number of the absorbing surface.

4.3.2 Sound pressure level of reverberated field

Sound pressure level of a ideal, diffuse sound field (statistical, uniform distribution), L_{diff} is depending on A_{equ} and the acoustic power level L_p and is calculated as

follows:

$$L_{diff} = L_p - 10log_{10}\frac{A_{equ}}{4m^2}dB. \tag{15}$$

4.3.3 Reverberation Time (RT)

One important criteria to describe the acoustics of a room is the reverberation time introduced by Sabine. The reverberation time measures the time, after a sound source in a room is turned off, until the energy density ω(t) decays to 1/1.000.000 part of its initial value or until the sound pressure decays to 1/1000 part, which corresponds to 60 dB, of the initial sound pressure. The RT is defined as follows:

$$T = 0,161[\frac{s}{m}] \cdot \frac{V}{A_{equ}} \tag{16}$$

where V is the volume of the room. As the desired decay of 60 dB can not always be achieved a decay of the sound pressure in a range between -5 dB and -35 dB is measured and doubled and results in the T_{30}. According to the frequency dependency of the absorption characteristics in a room, the reverberation time is frequency dependent as well.

4.3.4 Early Decay Time (EDT)

The EDT represents an improved measure of reverberance and envelopment. It considers the decay of the sound pressure level in a range between 0 dB and -10 dB, multiplied by the factor 6. The EDT is much shorter than the Sabine reverberation time. It is strongly affected by early reflections and thus is influenced by the room geometry and measurement position.

4.3.5 Critical distance

The critical distance is the distance to a sound source at which the energy density of the direct sound and the reverberant sound field are equal. The critical distance r_c for a point source with omni-directional sound radiation is

$$r_c = \sqrt{\frac{A_{equ}}{16\pi}}. \tag{17}$$

As many sound sources have a certain directivity pattern characterized by a certain gain factor g or directivity the critical distances is calculated by:

$$r_c = \sqrt{\frac{gA_{equ}}{16\pi}}. \tag{18}$$

5 Spatial hearing

An occuring sound event has a temporal and spatial representation. In this chapter the perception of sound in relation to the spatial representation of the sound event is described. The ear uses different attributes of the sound events to determine their position and spatial extensions. According to [Bla83] the attributes are divided into monaural and binaural signal attributes depending on whether one or both ears are required to detrmine the position and extension of the sound event. During the following considerations the head-related spherical coordinate system shown in figure 3 is used to discribe the position of sound events in relation to the head of the listener.

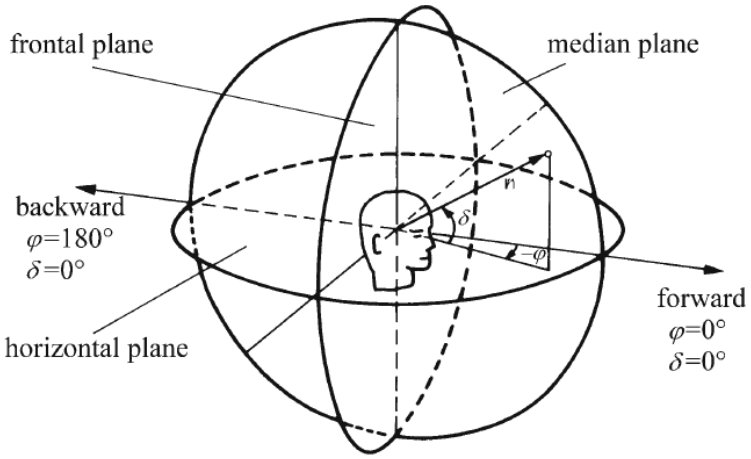


Figure 3: Head-related coordinate system [Bla83]

5.1 Head-related transfer functions (HRTFs)

HRTFs describe the impact of ears, head and torso on the perception of sound events related to their position, direction and frequency [Wei08, chap. 3, Blauert, Braasch]. The transmission path from the sound source to the eardrum is influenced by the outer ear, head and torso depending on the position, direction and frequency of the sound event. Amplitude and phase of the frequency spectrum are changed in connection to the specific position of the sound event. The transmission path and resulting spectral changes are described by a binaural pair of head-related transfer functions:

$$H_l(f, r, \varphi, \delta), H_r(f, r, \varphi, \delta) \tag{19}$$

where f represents the frequency, r denotes the radius to the sound source, φ is the azimuth angle and δ the elevation angle of the direction of sound incidence. Since the difference between both ear signals are required to resolve the spatial cues of the source signal, the HRTFs of the right and left ear are divided and result in the interaural transfer function:

$$H_i(f, r, \varphi, \delta) = \frac{H_r(f, r, \varphi, \delta)}{H_l(f, r, \varphi, \delta)} \tag{20}$$

5.2 Localization in the median plane

Sound signals arriving from the median plane cause very similar ear signals and thus are considered to have no interaural cues. A subjective evaluation [Bla83] investigating the localization in the median plane, using third octave band noise as stimulus presented at various directions, shows a localization dependency on the center frequency of the third octave band noise rather than on the actual position of the sound signal. The explanation of this phenomenon is based on the fact that due to the constitution of the outer ear certain spectral components of an arriving broad band signal are amplified and others attenuated, depending on the angle of incidence of the sound source. Thus the localization of the sound event at a certain angle of elevation depends on the presence of energy in certain frequency bands, so called directional bands.

5.3 Localization in the horizontal plane

With a sound incidence from the horizontal plane additional interaural cues contribute to the localization of the sound source: Interaural time differences (ITDs) and interaural level differences (ILDs).

5.3.1 Interaural time differences (ITD)

The interaural time differences $\Delta\tau$ are caused by the distance between the two ears, which means that there are arrival time differences between the two ear input signals. Blauert [Bla83] considers the maximum path difference for sound signals arriving at the two ears from the side with 21 cm. Thus the maximum ITD for a sound in the horizontal plane at an azimuth φ equal $\pm$ 90 is about 630 μs.

5.3.2 Interaural level differences (ILD)

Interaural level differences ΔL consider the level differences between both ear signals and can be described in terms of the amplitude of the interaural transfer function:

$$\Delta L = 20 \log |H_i(f, r, \varphi, \delta)| \tag{21}$$

Interaural level differences can lead to the lateral displacement of the sound event and depend strongly on the frequency of the source signal.

5.4 Perception of distance of a sound source

In the perceptual process of estimating the distance of a sound source multiple cues are employed by human hearing. Four “acoustic distance cues” are proposed by Zahorik [Zah02]: Sound intensity, direct-to-reverberant energy, spectrum and interaural time and level differences.

The **intensity** decreases with an increasing distance from the source to the receiver. Under ideal conditions of a point source in the acoustic free field, the intensity loss is a function of distance obeying the inverse-square law. In the case of reflective surfaces the intensity loss is decreased.

The **direct-to-reverberant ratio** decreases with the increase of the source distance. The change in direct-to-reverberant energy ratio is influenced by the impact of the inverse square law on the direct portion of the sound field. The later arriving reflected portion is approximated by the diffuse sound field, which has a uniform energy distribution over varying positions.

The **spectrum** changes for a distance greater than 15 m, according to Blauert [Bla83], are caused by the sound-absorbing properties of air. These are relatively small, around 3 to 4 dB loss per 100 meters at 4 kHz. A second type of spectral change occurs in reflective environments, which affects the spectrum due to the absorption properties of the reflective surfaces. With an increase of the distance of the source to the receiver the proportion of reflected energy increases and changes the at-the-ear spectrum systematically.

In the acoustic near-field the **binaural differences** in both, time and level are no longer independent of the radial distance to the sound source, as with planar waves in the far-field. Brungart and Rabinowitz [BDR99] investigated the localization of nearby sources and defined a region within 1 m distance to the listener as "proximal region". In the experiment an auditory point source was moved to a random position within 1 m of the subject's head, and the subject responded by pointing to the perceived location of the sound with an electromagnetic position sensor. The following observations can be summarized: (1) The ILD increase dramatically as the sound source approaches the head. This increase occurs even at low frequencies where head shadowing is negligible. (2) The ITD is roughly independent of distance in the proximal region. (3) The magnitude of the HRTF is relatively greater at low frequencies than at high frequencies when the source is near the head.

5.5 Multiple sound sources

In this chapter the spatial hearing in the sound field of two sound sources radiating coherent signals is described [Bla83].

5.5.1 Law of first wave front and perception of echo

Two loudspeakers in stereo arrangement, radiating non-periodic, coherent signals are considered. With simultaneous radiation of the signals with the same level the auditory event appears exactly in front. If one signal is delayed, with the delay increasing continuously from τ_{ph} equal 0, where τ_{ph} is the phase delay, the direction of the auditory event migrates to the position of the loudspeaker radiating the earlier signal. With a delay time 620 μs $\leq \tau_{ph} \leq$ 1 ms the auditory event will have reached the position of the loudspeaker radiating the earlier signal. With an increase beyond 1 ms, it can be observed that the direction of the auditory event remains nearly constant and the position is determined by the loudspeaker radiating the earlier signal. This phenomenon is called "law of the first wavefront".

This phenomenon is important for hearing in closed spaces and the signal radiated first is further considered as the direct sound source and the delayed signal as reflection. With the increasing delay beyond 1 ms the auditory event is louder and has a greater extent in space as if the direct sound is presented alone. With a further increase of the delay time the tone color of the auditory event changes and its spatial

extent further increases. When a certain delay time is exceeded the auditory event separates in two events in different directions. The direction of the first event is determined by the primary sound and generally agrees with the direction of incidence of the primary sound. The second auditory event appears in the direction of incidence of the reflection, called echo. The shortest delay time at which the second auditory event becomes audible is the "echo threshold". The echo threshold represents the upper limit of validity for the law of the first wavefront. At delay times less than 50 ms, echoes are no longer perceived as annoying even if the reflection is considerably stronger than the primary sound, which is known as "Haas effect".

The perception of an echo depends on the type of the signals presented, on the delay time of the reflection and on the level difference between direct sound source and reflection. The echo threshold is greater for signals with long duration compared to pulse signals and high-frequency signal components lead to shorter delay times compared to low-frequency components. If the level of the reflection is raised the echo appears at a shorter delay time and reciprocal a longer delay time occurs if the reflection level is lowered. Depending on the signal level the smallest echo thresholds are derived for "clicks" with less than 2 ms. For speech the threshold is about 20 ms. Pulsed sinusoidal signals with long duration result in a echo threshold of 180 ms [Bla83, p. 230-235].

6 Subjective perception in closed spaces

In this chapter the subjective perception in closed spaces are described. There is multitude of investigations in this field and a difference in the psychoacoustic terminology in the description of the subjective perceptions. In the following discussion relevant parameters and perceptions to describe the sound pattern of the instrument are described.

6.1 Sensation of space

The "'sensation of space"' is caused by sound reflections reaching the listener from different directions [Kut09, chap. 7.7]. As the human hearing cannot locate these directions separately it processes them into an overall impression, the sensation of space. Spaciousness is not primarily caused by a stationary directional distribution of reflections. Instead quite few synthetic reflections, which reach the listener from lateral directions may cause the sensation of spaciousness. Thus the reflections are assumed to carry mutually incoherent signals. Lateral reflections, with a delay of 5 to 80 ms contribute to spatial impression, depending on their energy and $\cos\varphi$, where φ is the angle between the axis through the ears of the listener and the direction of sound incidence. In addition the contribution of the lateral reflections to spaciousness is independent from the presence or absence of reverberation. Different authors describe this sensation with quite different terminologies. Griesinger [Gri97] gives an overview of the psychoacoustic terminology and uses the terms spaciousness and envelopment as synonymous to describe as the perception occuring within a large concert space [Gri99]. Vorlnder [Vor08, chap. 6.4.4] summarizes the separation of spaciousness into two kinds of spatial impressions, the "apparent source width" (ASW) and "listener envelopment" (LEV) which are described in the following chapters.

6.1.1 Apparent source width (ASW)

The early part of the lateral reflections, up to 80 ms contributes to the ASW. For the source localization in complex sound fields the human hearing determines the direction of the sound incidence by the first arriving sound event. Early lateral reflections thus add some uncertainty the localization of the sound source direction. This leads to the localization of an extended source with a certain width (ASW). The objective qualities to evaluate ASW are the Lateral Fraction (LF), described in chapter 7.1 and the interaural cross correlation (IACC), described in chapter 7.3.

6.1.2 Listener envelopment LEV

The spatial impression of the late lateral reflections is termed listener envelopment (LEV). Investigation results of Soulodre et al. [SLC03] confirm the results of previous investigations which show that the LEV is primarily determined by the level and spatial distribution of the late energy. Secondary factors influencing the LEV are the overall playback level and the reverberation time. The objective measure LG_{80}^{∞} which investigates the strength of the lateral reflections correlates with the prediction of LEV and is described in detail in chapter 7.2.

6.2 Perceptibility of single reflections

In the case of only one or a few reflections of a direct signal within a room, two main aspects are of importance, the general perceptibility of single reflections and the impact of the reflections on the listening impression.

In order to receive a threshold level to determine weather a reflection is audible or completely masked by the direct sound source according to [Kut09] the "audibility threshold" is defined as follows:

$$\Delta L = -0.575t_0 - 6dB \tag{22}$$

where ΔL is the relation of reflection pressure level to direct sound pressure level, with t_0 representing the time delay in ms. To find this threshold two configurations of a sound field, one with the presence of specified reflections, the other without sound reflection, are presented to the subject. The subject is asked weather a difference is noticed or not. A level at which 50% of answers are positive is defined as audibility threshold.

In the case of the perception of a reflection an echo, as already mentioned in chapter 5.5.1, may be percepted. Another perception is the perception of "coloration". Coloration is caused by the superposition of strong, isolated reflections and the direct signal or in case of a sequence of equidistant reflections. The superposition of the direct signal and a its delayed version results in comb filter structure which is explained in chapter 8.4.1. The absolute threshold of audible coloration caused by a comb filter rises with an increase in the delay time.

Griesinger [Gri97] summarizes the results of previous observation in localization and spatial impression of single reflections. With a band filtered noise as frontal direct sound source, an enveloping spatial impression is distinctly audible even if the reflection is 20 dB weaker, compared to the frontal direct sound. For frequencies below 700 Hz the audibility threshold rises with the reflection angle converging to the direct source position. With a single noise-like signal there is very little change in the spatial properties of the sound as the delay of the reflection is increased beyond 10 ms. The reflection cannot be determined unless it is within 3 dB difference to the direct sound. With an increasing reflection level from -20 dB the impression of envelopment increases while the perception of the source stays sharp. With exceeding -3 dB the apparent source width increases and the source position move towards the reflection.

In his investigations Griesinger used a band filtered noise below 300 Hz and found that a longer delay, compared to previous investigations, is required for the surround effect. With sufficient delay the ASW decreases and the surround perception is induced. Six lateral reflections of equal amplitude (-10dB), with a randomly spread delay between 5 ms and 50 ms, did not produce a surround effect below 2000 Hz. With adding the delayed sound components to the direct sound, an increase of the source width was noticed. With the spread of the delay between 40 ms to 100 ms the source stayed able to localize and the sense of surround was was identical to higher frequencies. With a delay of 50 to 100 ms, which are considered to be sufficient to give a surround effect with noise, the reflections were heard individually and not as single source in the presence of a surround sound.

7 Objective measures of subjective perception

7.1 Lateral energy fraction (LF)

[Wei08] A quantity measuring the spatial impression, commonly used in concert-hall acoustics is the lateral energy fraction LF, which relates the lateral sound energy to the total sound energy in the first 80 ms of an impulse response. The calculation of the LF is quoted in equation 23.

$$LF = 10log_{10}\left[\frac{\int_0^{80ms} p_F(t)^2 dt}{\int_0^{80ms} p_o(t)^2 dt}\right][dB] \tag{23}$$

where $p_F(t)$ is the instantaneous sound pressure derived by using a figure-of-eight microphone and $p_o(t)$ is derived by an omnidirectional microphone. The derived LF value is associated to the apparent source width, since with an increasing value of LF the apparent width of the source increases. Desirable values of the LF are specified with 0,10 < LF < 0,25 [Wei08].

7.2 Lateral hall gain (LG)

In the investigation of listener envelopment (LEV), Soulodre et al. [SLC03] proposed the objective measure LG_{80}^{∞}, which is related to the sum of lateral energy arriving after 80 ms. The investigations show a high correlation of the LG_{80}^{∞} with the subjective perception of LEV. LG_{80}^{∞} is defined as:

$$LG_{80}^{\infty} = [\frac{\int_{0.08}^{\infty} p_F^2(t) dt}{\int_0^{\infty} p_A^2(t) dt}][dB] \tag{24}$$

where $p_F(t)$ is the lateral sound pressure, measured in the experiment, by using a figure-of-eight microphone and p_A is the response at a distance of 10 m in a free field.

7.3 Spatial binaural factors according to Ando

Ando [And09] defines the interaural cross correlation function (IACF) based on the binaural impulse response (BRIR) $h_L(t)$ and $h_R(t)$. The maximum value of the IACF is the interaural cross correlation coefficient (IACC) and is related to the spatial impression of a sound field. The normalized interaural cross correlation function is defined by:

$$\Phi_{lr}(\tau) = \frac{\int_{\Phi_{lr}(\tau)}}{\sqrt{\Phi_{ll}(0)\Phi_{rr}(0)}} \tag{25}$$

where $\Phi_{ll}(0)$ and $\Phi_{rr}(0)$ are the ACFs for τ equal 0 for the left and right ears and τ is the interaural time delay within plus and minus 1 ms. The equation can further be written as [DGG09]:

$$\Phi_{lr}(\tau) = \frac{\frac{1}{2T}\int_{-T}^{T} h_L(t)h_R(t+\tau)dt}{\frac{1}{2T}\sqrt{\int_{-T}^{T}[h_L(t)]^2 dt \int_{-T}^{T}[h_R(t)]^2 dt}} \tag{26}$$

The interaural cross correlation coefficient IACC is calculated from the IACF(τ) as follows:

$$IACC = max|\Phi_{lr}(\tau)| \tag{27}$$

$$|\tau| \leq 1ms \tag{28}$$

Based on the IACF Ando defines 3 spatial factors, which are illustrated in figure 4:

- The interaural correlation magnitude IACC is the maximum value of the IACF, and is associated with the subjective diffuseness of the sound.
- The interaural delay τ_{IACC} is the delay at which the IACF has its maximum value of IACC, and is associated with sound direction in the horizontal plane.
- The W_{IACC} is the width of the maximal IACF peak, defined by the size of the delay range over which the IACF peak is at least 90 % of its maximal value (δ = 0.1*IACC).

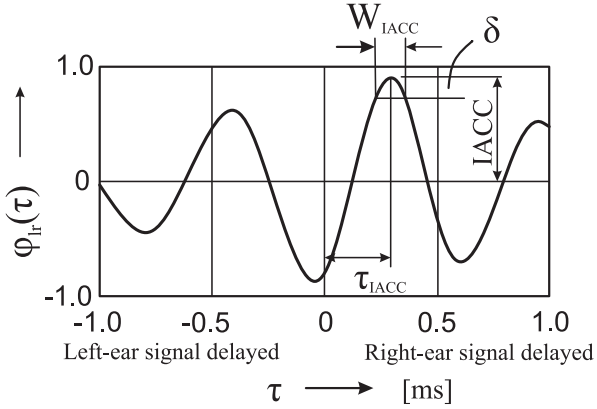


Figure 4: Definition of the three spatial factors extracted from the interaural correlation function (IACF) [And09].

According to the descriptions of Ando [And09] with an observed τ_{IACC} equal 0, the preferred condition of a frontal sound image and a well-balanced sound field are perceived. The width of the IACF, given by the W_{IACC} is associated with the apparent source width (ASW). The ASW may be perceived as a directional range corresponding mainly with the W_{IACC}. A well-defined directional impression corresponding to the interaural time delay τ_{IACC} is perceived with listening to a sound with a sharp peak in the IACF which implies a small value of W_{IACC}. With a low value for the IACC <0.15 a subjectively diffuse sound is perceived. Thus a minimal value of IACC corresponds to maximal diffuseness and listener envelopment (LEV).

The choice of the time limits between the coherence of both ear signals is investigated, the frequency filtering and the subjective evaluation is not standardized

[Wei08, chap. 5]. The time limits are selected in terms of a separate investigation of the early reflections and late reverberation. The frequency should in general be filtered in octave bands between 125 Hz and 4000 Hz and the integration limits may be chosen as follows:

- $IACC_{E(arly)}$: $t_1 = 0$ ms; $t_2 = 80$ ms
- $IACC_{L(ate)}$: $t_1 = 80$ ms; $t_2 = 500....2000$ ms
- $IACC_{A(ll)}$: $t_1 = 0$ ms; $t_2 = 500....2000$ ms

The selected frequency range is indicated by an additional index, e.g. $IACC_{E3B}$ for $IACC_E$ which indicates an averaging of three octave-frequency ranges 500, 1000 and 2000 Hz with $t_1 = 0$ ms and $t_2 = 80$ ms.

8 Room acoustic modeling methods

In this chapter room acoustic modeling methods with regard to a real-time implementation are discussed. As described in chapter 3 the methods can be divided into wave-based, geometric-based and statistics-based methods.

8.1 Wave-based methods

Wave-based methods aim at numerical solution of the wave equation, which is computationally more expensive and real-time processing is only practicable for low frequencies and small room volumes. Wave-base methods are the boundary-element method, BEM, and the finite-element method, FEM [KDS93]. The FEM allows an easy determination of the cavity eigenfrequencies. The computational requirements of both techniques increases very rapidly with frequency and thus limits their real-time approach to low frequencies and small rooms.

8.2 Geometrical-acoustics-based methods

Geometric-based methods rely on geometrical room acoustics [Kut09, chap. 4]. In geometrical room acoustics the concept of waves is replaced by the concept of sound rays, thus a number of simplifications and idealizations, which are discussed in chapter As described in chapter 4.2, can be assumed. These simplifications allow a considerably faster calculation compared to the wave-based methods but interference and diffraction are not easily considered with this method. Two main approaches to determine the early reflection paths from the virtual sound source to the receiver position are distinguished: The *ray-tracing* and the *image source* method:

8.2.1 Ray tracing

[Vor08]In the ray-tracing method the sound source is assumed to radiate sound rays. Each ray carries a certain energy and propagates with the speed of sound. The rays are followed through the domain until they get attenuated below a certain level, due to multiple reflection, they leave the domain, or reach the listening position. In most of the cases the reflections are modeled as specular reflections. The modeling of diffuse reflections is possible but at the cost of extra computation.

8.2.2 Beam tracing

[FCE+98] Beam tracing methods classify reflection paths of a source by recursively tracing pyramidal beams, which are represented by a bundle of rays. Briefly, a set of pyramidal beams, which cover the 2D space of directions from the source, are constructed. In the case of the detection of intersecting polygones the original beam is clipped and a reflection beam is contructed by mirroring the transmission beam over the intersected polygone. The beam paths are stored in a beam-tree data base which is updated in during the run-time. An computationally efficient approach for real-time auralization of complex geometries employing an accelerated beam tracing algorithm is proposed by Noisternig et al. [NKSS08].

8.2.3 Image source method

In the image source method the reflection paths from the source to the listener are calculated by sequentially mirroring the sound source at the room boundaries. The image source principle is described in chapter 4.2.1. Berkley et al. [AB79] proposes an image model for the room acoustic modeling in simple rectangular enclosures. The extension of the model [Bor84] provides the calculation of the image source model for arbitrary polyhedral geometries.

8.3 Statistics-based methods

Due to calculation time constraints the early part of the room impulse response is computed in detail while the late reverberation part of the impulse response is computed by a statistical model. Statistics-based models assume a statistical distribution of sound energy and is thus based on statistical room acoustics, as described in chapter 4.3.

8.4 Reverberation models

In this chapter the real-time generation of natural and realistic sounding synthetic reverberation is discussed. Subsequent reverberation can be statistically described and is not depending on the source and receiver position, thus it leads to a characterization of the room itself. Challenges in the design of artificial reverberators are to produce sufficient density of modes and echoes in the frequency domain and time domain and to avoid unpleasant resonances especially to transient signals. Furthermore the reverberation time is supposed to decrease as function of frequency in order to simulate the low-pass-filtering characteristics of air-absorption and materials. The basic elements to simulate exponentially decaying room impulse responses are the recursive comb filter and the all-pass filter based on the Schroeder Algorithm [Sch61].

8.4.1 Comb filter - all-pass filter

Compared to the recursive comb filter the magnitude response of an all-pass filter equals one across the whole frequency range. This property eliminates the strong coloration caused by an comb filter with stationary input signals but not with transient signals. There are two approaches with the above filter types to improve the echo density:

- The parallel comb filter increases echo density but cannot achieve a flat frequency response.
- The series all-pass filter results in a new all pass filter with flat frequency response but coloration with transient signals cannot be avoided.

The number of eigenfrequencies per Hz, at a frequency f is an important parameter to characterize subsequent reverberation [Kut09]. The density of modes increases with the frequency independant from the room shape. The frequency density f_d quotes as follows:

$$f_d = \frac{4\pi V}{c^3} \cdot f^2. \tag{29}$$

The density of echoes in time domain increases with the square of time. The number of reflections per second N_{ref} quotes as follows:

$$N_{ref} = \frac{4\pi c^3}{V} \cdot t^2. \tag{30}$$

After a certain period of time t_c reflections overlap and show statistical delay behavior:

$$t_c = 5 \cdot 10^{-5} \sqrt{\frac{V}{\Delta t}} \tag{31}$$

where Δt is the pulse width of the excitation signal. With the parallel comb filter method by Schroeder, insufficient echo density within the room response is derived and thus the general delay network according to [JCry] is proposed.

8.4.2 General Delay Network

According to investigations of Jot [JCry] a frequency-dependent decay time for different frequencies can be derived controlling the pole positions of the FDN structure. Jot further proposes that all modes in a narrow frequency band should decay at the same rate to avoid unpleasant resonances and isolated "ringing modes" at the end of the reverberation response. The general delay network, illustrated in figure 5 contains a number of delay lines whose output signals are feed back to the inputs of the delay lines via a "feedback matrix".

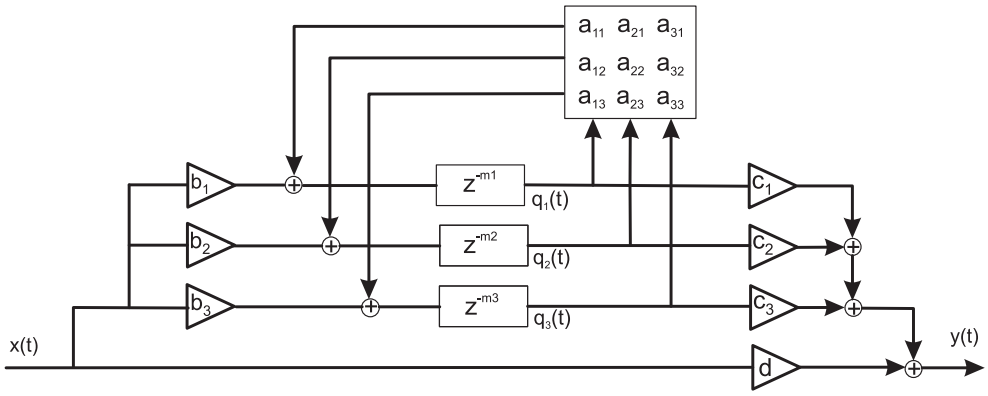


Figure 5: Feedback delay network [JCry]

According to figure 5 the FDN comprises a number of N delay lines, each τ equal m_i T_s seconds long, where T_s equal $1/f_s$ is the sampling interval. The FDN is completely described by the following equations [Zöl02]:

$$y(n) = \sum_{i=1}^{N} c_i s_i(n) + dx(n) \tag{32}$$

$$s_i(n + m_i) = \sum a_{i,j} s_j(n) + b_i x(n) \tag{33}$$

where s_i(n), 1 ≤ i ≤ N, represent the delay outputs at the n-th time sample. If m_i equal 1 for every i, the state space description of a discrete-time linear system

is derived. The number of the order of a FDN is indicated by m_i. From the state-variable description the system transfer function of a FDN is derived:

$$H(z) = \frac{Y(z)}{X(z)} = c^T[D(z^{-1} - A]^{-1}b + d \tag{34}$$

where D(z) = diag(z^{-m1},...,z^{-mN}) is the delay matrix and A=$[a_{i,j}]_{NxN}$ is the feedback matrix. The stability of the system depends on the properties of the feedback matrix A. The fact that $\|A\|^n$ decays exponentially with n ensures the stability of the FDN structure. The poles of the structure are obtained by the solutions of:

$$det[A - D(z^-1)] = 0. \tag{35}$$

According to [Jot97] a "prototype network" is defined as network having only non-decaying and non-increasing eigenmodes. This condition is only valid if all system poles have unit magnitude, which further imposes the system to be loss-less and with infinite reverberation time. A unitary feedback matrix is suited to fulfill this condition. (An unitary network is supposed to be the multichannel equivalent of an all-pass filter structure.) [Zöl02] In order to obtain an exponential decaying impulse response every delay unit is attenuated by:

$$g_i = \alpha^{mi}. \tag{36}$$

This corresponds to a multiplication of all poles with the same factor α, which imposes all modes to decay with the same rate. A reverberation time for an attenuation level of 60 dB is derived by:

$$T_d = \frac{-3T_s}{log_{10}\alpha}. \tag{37}$$

A frequency-dependent decay time is now derived by introducing frequency-dependent attenuation $g_i(\omega)$ using so called "absorptive filters":

$$20log_{10}|g_i(\omega)| = \frac{-60\tau_i}{Tr(\omega)} \tag{38}$$

where $\tau_i = m_i$ T is the delay length in seconds and T represents the sample period. For the practical realization of the prototype network the most important question is how to design the feedback matrix A.

- Matrix A should have no coefficients identical to 0, to receive a fast increase in the echo density with circulation through multiple delay lines.
- The crest factor of the matrix A, which is the ratio of largest coefficient over RMS average of all coefficients, indicates the speed of convergence towards Gaussian amplitude distribution and should be minimized.

Several families of unitary matrices can satisfy this criterions and allow to minimize the complexity of an implementation on a programmable processor. A comparison of unitary matrices is provided by Jot [Jot97] and the Housholder matrix is of the type:

$$A = (\frac{2}{N}) \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} [1...1] - I. \tag{39}$$

Housholder matrices involve only 2N operations but have a high crest factor for large values of N, which can be avoided by dividing the system into smaller unitary systems.

9 Sound spatialization

In the implementation of auditory virtual environment approaches an accurate reproduction of the sound field fundamentally contributes to the quality of the perceived environment. In addition the evocation of auditory events located all over a three-dimensional plane around the listener is desired. As mentioned in chapter 3 methods for direct spatial encoding of the room impulse response include e.g. Vector Base Amplitude Panning (VBAP) [Pul97] and direct convolution of the impulse response with HRTFs. Higher Order Ambisonics (HOA) [Dan00], [Ger73] and Wave Field Synthesis (WFS) [SRA08] are feasible sound field reproduction techniques which aim at the accurate reproduction of the sound field.

HOA and WFS aim at the physical reconstruction of the sound field and provide exact solution of the sound wave equation. In both cases sound recording and reproduction by respective microphone and loudspeaker arrays enables the acoustic encoding and decoding of the spatial sound information. A further investigation of both techniques is given by Daniel [DNM03]. In the following the theoretical basement of both approaches is explained. Only a brief description of the WFS is provided as the as the sound spatialization in the practical implementation of the instrument is derived using HOA.

9.1 Kirchhoff-Helmholtz integral

[SRA08] A loudspeaker array surrounding the listener can be regarded as an inhomogeneous boundary condition for the wave equation. The solution of the homogeneous wave equation for a bounded volume V, with regard to the inhomogeneous boundary conditions is based on the Kirchhoff-Helmholtz integral (KHI). The KHI states that a pressure field P(x,ω) inside a source free volume V, caused by an arbitrary sound source distribution, can be described by the pressure P(x_0, ω) along the surface ∂ V enclosing the volume and the gradient of the pressure normal to the surface $\frac{\partial}{\partial n}$ P(x_0, ω). Thus the KHI writes as:

$$P(x,\omega) = -\oint_{\partial V} (G(x|x_0,\omega)\frac{\partial}{\partial n}P(x_0,\omega) - P(x_0,\omega)\frac{\partial}{\partial n}G(x|x_0,\omega))dS_0 \qquad (40)$$

where (G(x—x_0, ω) denotes a suitable chosen Green's function. Assuming free-field propagation within V the Green's function is the free-field solution of the wave equation an is called *free-field Green's function* $G_0(x|x_0,\omega)$. If the Green's function is realized by a continuous distribution of monopole and dipole sources placed on the boundary ∂ V, the wave field is fully described. This fact is used in the reproduction of the sound field via loudspeakers, representing the secondary sound source distribution. In the practical approach the dipole sources are removed, since loudspeakers are capable of reasonably reproducing monopole sound sources. Higher Order Ambisonics and Wave Field Synthesis apply two different techniques for the elimination of the dipole part. In WFS the second term of the Kirchhoff-Helmholtz integral in equation 40, which represents the the dipole secondary source distribution, is discarded by the using the so called Neumann Green's function. A detailed description of the WFS theory is given in [SRA08].

9.2 Higher Order Ambisonics (HOA)

Higher Order Ambisonics is an extension of the Ambisonics approach, mainly introduced by Gerzon [Ger73], to derive a higher spatial resolution in the reproduction of the sound field. In the following a short description of the deviation of the encoding and decoding of the sound field is given, according to [Dan00], [DNM03]. HOA is based on the decomposition of the sound field into a series of spherical harmonic functions, which can be truncated at an arbitrary order. Using higher orders contribute to the improvement of the accuracy in the reproduction of the sound field and leads to the enlargement of the sweet spot to a sweet listening area.

9.3 Mathematic fundamentals

The Ambisonics representation is based on the solution of the wave equation in spherical coordinates for the central listening point at $\vec{r}$=0. In Ambisonics the reproduced, virtual sound sources and loudspeakers are assumed to emit plane waves, assuming far field sources. The vector $\vec{r}$ is defined by the radius r, azimuth φ, elevation δ. The denomination of the angles is conform to the head-related spherical coordinate system in chapter 5, figure 3.

The pressure field can be written as:

$$p(\vec{r}) = \sum_{m=0}^{\infty} j^m j_m(kr) \sum_{0\leq n\leq m, \sigma=\pm 1} B_{mn}^{\sigma} Y_{mn}^{\sigma}(\varphi, \delta) \tag{41}$$

with the sound pressure p and the wave number k=2πf/c, the spherical Bessel functions $j_m(kr)$ and the spherical harmonic functions Y_{mn}^{σ}. For the center of the listening area, free of virtual sources only the "through-going" field according to Daniel [DNM03] is considered and the spherical Hankel functions are considered being equal 0. The resulting sound field is built up by a superposition of spherical harmonic functions. As in practice only a finite number of components can be reproduced the series is reduced to the order M, which is the Ambisonics order:

$$p(\vec{r}) = \sum_{m=0}^{M} i^m j_m(kr) \sum_{0\leq n\leq m, \sigma=\pm 1} B_{mn}^{\sigma} Y_{mn}^{\sigma}(\varphi, \delta). \tag{42}$$

The spherical harmonic functions $Y_{mn}^{\sigma}(\varphi, \delta)$ consist of the Legendre functions multiplied by sine and cosine terms. Each order m consists of 0≤n≤m spherical harmonic functions with values of σ=±1.

$$Y_{mn}^{\sigma}(\varphi, \delta) = \bar{P}_{mn}(sin\delta) \cdot \begin{cases} \cos(n\varphi) & \text{if } \sigma = +1 \\ \sin(n\varphi) & \text{if } \sigma = -1 \end{cases} \tag{43}$$

where $\bar{P}_{mn}(\varsigma)$ is the semi-normalized Legendre polynomial (semi-orthonormalization by Schmidt).

$$\bar{P}_{mn}(\varsigma) = \sqrt{\epsilon_n \frac{(m-n)!}{(m+n)!}} P_{mn}(\varsigma) \tag{44}$$

with ϵ_0=1 and ϵ_n=2 for n≥1. Figure 6 illustrates a 3D view of spherical harmonic functions according to Daniel [DNM03].

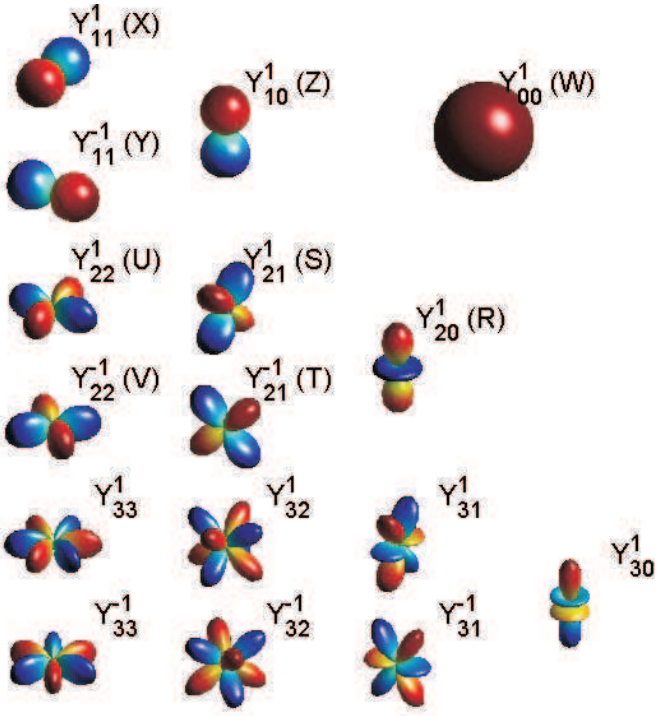


Figure 6: 3D view of spherical harmonic functions with usual designation of associated Ambisonics components [DNM03]

9.4 Directional encoding equations

The spherical harmonic decomposition of a plane wave of incidence (φ_S, δ_S) conveying the signal S leads to the expression of the Ambisonics components [DNM03]:

$$B^{\sigma}_{mn} = S \cdot Y^{\sigma}_{mn}(\varphi_S, \delta_S). \tag{45}$$

Thus a far field source is encoded by applying real gains to the received pressure signal S. The number N of Ambisonics channels is depending on the Ambisonics order and is in the 2D case defined by $N_{2D} = 2\text{m}+1$. In a 3D Ambisonics system the channel number is derived by

$$N_{3D} = (M+1)^2. \tag{46}$$

The Ambisonics channels are termed according to Daniel [Dan00] with B^{σ}_{mn}:

$$\begin{bmatrix} B^{+1}_{00} \\ B^{+1}_{11} \\ B^{-1}_{11} \\ B^{+1}_{10} \\ B^{+1}_{22} \\ B^{-1}_{22} \\ \vdots \end{bmatrix} \tag{47}$$

9.4.1 Encoding functions

A precise definition of normalization schemes is given by Daniel [Dan00]. The following table shows semi-normalized, three-dimensional (SN3D) encoding functions until the 2^{nd} order.

Order	${B_{mn}}^{\sigma}$	$({}_{mn}{}^{\sigma})$	$(Y_{mn}^{\sigma)(SN3D)}(\varphi,\delta)$
0	W	$\binom{1}{00}$	1
	X	$\binom{1}{11}$	cos φ cos δ
	Y	$\binom{-1}{11}$	sin φ cos δ
1	Z	$\binom{1}{10}$	sin δ
	U	$\binom{1}{22}$	$\sqrt{3/2}$cos(2φ)cos^2 δ
	V	$\binom{-1}{22}$	$\sqrt{3/2}$sin(2φ)cos^2 δ
2	(S)	$\binom{1}{21}$	$\sqrt{3/2}$cos φsin(2δ)
	(T)	$\binom{-1}{21}$	$\sqrt{3/2}$sin φsin(2δ)
	(R)	$\binom{1}{20}$	(3 sin^2 δ - 1)/2

Table 1: Higher Order Ambisonics SN3D encoding functions until 2. order [Dan00, p.151]

9.5 Decoding of the sound field

The Ambisonics decoding is based on the "re-composing principle" proposed by Daniel [Dan00] which aims at recomposing the encoded Ambisonics components $\bar{B}_{mn}^{\sigma}$ in the center of the speaker array. As the loudspeaker distance to the listening point is assumed to be high enough their signals S_i are encoded as plane waves with the coefficient vectors c_i [DNM03]:

$$c_i = \begin{bmatrix} Y_{00}^{+1}(\varphi_i,\delta_i) \\ Y_{11}^{+1}(\varphi_i,\delta_i) \\ Y_{11}^{-1}(\varphi_i,\delta_i) \\ \vdots \\ Y_{mn}^{\sigma}(\varphi_i,\delta_i) \\ \vdots \end{bmatrix} \bar{B} = \begin{bmatrix} \bar{B}_{00}^{+1} \\ \bar{B}_{11}^{+1} \\ \bar{B}_{11}^{-1} \\ \vdots \\ \bar{B}_{mn}^{\sigma} \\ \vdots \end{bmatrix} S = \begin{bmatrix} S_1 \\ S_2 \\ S_3 \\ \vdots \\ S_N \end{bmatrix} \tag{48}$$

The re-encoding equation written in matrix notation, with the "re-encoding matrix" C=[c_1 .. c_N] quotes as:

$$\bar{B} = C \cdot S. \tag{49}$$

The decoding process aims at deriving the loudspeaker signals S from the original Ambisonics signals B:

$$S = D \cdot B \tag{50}$$

where D is the decoding matrix. In order to ensure that $\bar{B}$=B, the system 49 must be inverted. Thus the decoding matrix D is defined as:

$$D = pinv(C) = C^T \cdot (C \cdot C^T)^{-1}, \tag{51}$$

provided that there are enough loudspeakers, e.g. N$\geq N_{2D}$ or N$\geq N_{3D}$.

9.6 3D Ambisonics based binaural sound reproduction

In conventional binaural sound reproduction systems spatialization is performed by convolving the source signal with HRTFs. This approach requires a time-varying interpolation between different HRTFs which yields the problem of artifacts decreasing the localization performance of the system. In order to overcome this problem the usage of a virtual Ambisonics approach is proposed [NSMH03]. This approach is based on the idea of decoding ambisonic to virtual loudspeakers. The binaural signals are derived by convolving the virtual loudspeaker signals with HRTFs according to their position in space. The right and left ear signal are derived by the superimposition of the filtered signals. Thus the virtual Ambisonics approach results in the usage of a bank of time-invariant HRTF filters. The actual head rotation is determined by the use of head-tracking and which is further taken into account by rotation matrices in the Ambisonics domain.

10 Instrumental interface

Within this chapter the structure of a tabletop tangible user interface (TUI) with regard to an application as musical interface is described. The term digital musical instrument and basic requirements concerning temporal aspects in the performance of music are defined. The concept of tangible interaction in combination with direct-touch interaction is proposed. Basic considerations concerning playability, apprenticeship and efficiency of a digital musical instrument are discussed.

10.1 Digital Musical Instrument

A digital musical instrument (DMI) is described [WD04] by Wanderley as computer-based musical instrument consisting of an gestural interface seperated from the sound generation unit processing the instrument output in real-time. Control unit and sound generation are related through a mapping strategy. Performer gestures are considered as all actions produced by the instrumentalist during the performance. The gestural controller is defined as input device representing the physical interaction interface of the instrument. Jordà [Jor02] proposes to consider a digital musical instrument as whole concept, which means control unit and sound generation should be merged within one structure. Thus the communication between players and their musical output should be improved.

10.2 Timing of interaction

Especially in the context of music time plays a central role. The possibility of precise timing of a musical gesture, e.g. triggering a musical sequence, a sonic reaction in real-time and the possibility of synchronization of processes are basic requirements in the development of an instrument. Mäki-Patola et al. observed in an experimental study of human latency tolerance for gestural sound control [MPH04b] a just noticeable difference (JND) threshold of 30 milliseconds for the perception of latency in playing a continuous sound instrument. 16 musically trained subjects played the Teremin and in each playing task, each example was to be reproduced at two different latency settings, one setting always was of zero latency. In the comparison of both settings the subjects were asked to evaluate the which of the two settings had larger latency. In an additional experiment [MPH04a] the influence of latency on the playing accuracy was investigated by measuring the time to reach desired notes on the Theremin. The study suggested that latencies up to 60 ms do not impair playing the instrument. An investigation of the performance of percussion instruments [DB01] showed that latencies over 55 ms degrade the playing performance. Four musical trained subjects were playing a baton instrument along with a metronome while latency was increased in small steps. With a delayed feedback of more of 55 ms the subjects showed increased difficulties in maintaining a steady rhythm.

10.3 Tabletop interaction

In current digital instrument design various control and interaction methods, beyond usage of mouse or keyboard as control units are investigated in order to adapt human-computer interaction processes to human perception abilities. Auditory and visual perception as well as the tactile sense are exploited to improve the bi-directional communication between player and instrument.

In tabletop systems the table surface serves as input interface and as display device to visualize the current instrument state and performing processes. The basic components of the structure include the acquisition of graphical data, object tracking and the interpretation of the tracking data. The derived object events are transported via transport protocol layer to a client application.

Object tracking is divided into three steps. (a) object identification, (b) frame by frame motion capture, (c) analysis of motion path to determine tendency in motion.

Depending on the applied input objects the interaction is categorized into:

- tangible interaction
- direct-touch interaction
- stylus interaction

An important criterion determining the quality of tabletop interaction is the number of objects which can be identified and reliably be tracked. Tabletop-frameworks which support the setup of the interactive graphical surface of tabletop interfaces are e.g. Anoto SDK [Gro96], Touchlib [Gro], Diamond-Touch SDK [Lab] and Reactivision, which is explained in chapter 11.3. The Anoto SDK enables stylus interaction, Touchlib and Diamond-Touch SDK support the implementation of direct-touch surfaces and Reactivision is a mix-framework which enables both tangible and direct-touch interaction. With regard to instrumental interaction the use of a mix-framework is beneficial because a high number of distinguishable input gestures is derived. In the following chapters these two interaction methods are discussed.

10.3.1 Concept of tangible user interface (TUI)

In 1995 the concept of "graspable user interfaces" was introduced by Fitzmaurice, Ishii, and Buxton [FIB95]. This concept was further extended by Ishii and Ullmer at the Massachusetts Institute of Technology and renamed into "tangible user interface" (TUI). In Ullmer [Ull95] a TUI is considered as system which uses spatially reconfigurable physical artifacts, so-called "tangibles", ("tangible" originates from the Latin "tangibilis" and "tangere", and means "to touch".), as representations and controls for digital information.

The physical properties of tangibles are categorized as follows:

- physically embodied
- physically representational
- physically manipulable
- spatially reconfigurable

Physically embodied implies a physical representation of digital information or functionality. As every computational device has a certain physical input possibility the focus in this context lies on the interaction modality with digital information and how people perceive the process. In contrast to physically embodied, "intangible" visual or auditory modalities are not physically accessible to direct haptic manipulation.

The term **physically representational** describes the association of particular digital functionality with artifacts or physical representations. There is a wide range of physical representations and TUI artifacts may be literally, iconically or symbolically representational. In order to provide a distinction possibility between multiple digital functionality it is important to consider different forms of physical artifacts.

In addition to physical representation the **physical manipulation** ability of interface artifacts represents the physical control capacity of the interface. An important fact for the accessibility of digital information is that the artifacts are graspable, which means they can be taken within the hand.

Spatially reconfigurable implies a variable positioning of the physical control mechanisms. In interaction with a tangible user interfaces the spatial reconfiguration is of central importance and implies the placement and removal, translation and rotation of the physical artifacts.

According to above explanation the key-characteristics of tangible interaction can be summarized as follows:

- Tangible representations are computationally coupled to underlying digital information
- Tangible representations embody mechanisms for interactive control
- Tangible representations are perceptually coupled to intangible representations
- The physical state of interface artifacts partially embodies the digital state of the system.

10.3.2 Direct-touch interaction

In direct-touch interaction an indirect mouse or pointing device is substituted by enabling direct-touch manipulation of graphical elements presented on the tabletop display. Depending on the structure of the framework uni-manual or bi-manual interaction with the implementation of single or multi-finger input is distinguished. The direct-touch input is further differentiated in terms of discrete or continuous interaction gestures. A discrete gesture enables pointing on a restricted area on the touch-screen whereas continuous interaction supports drawing a continuous line with the finger. As mentioned in chapter 10.2 the timing of musical processes is a crucial factor. In investigations of direct-touch and mouse input [FWSB07] in terms of execution time necessary to complete uni-manual and bi-manual tasks, the results show benefits of direct-touch interaction, in case of bi-manual tasks. Investigations [KAD09] comparing four input methods, in terms of execution time of multi-target selection show about twice as fast multi-touch selection compared to mouse-based selection. In particular the four methods, one mouse (indirect, uni-manual), one finger (direct-touch, uni-manual), two fingers, one on each hand (direct-touch, bi-manual), and any number of fingers (direct-touch, bi-manual, multi-finger) were compared.

Music making in the traditional sense is a bi-manual and multi-finger activity. Thus in the development of a digital musical instrument the possibility of controling processes by the use of direct-touch is desireable, as it enables a more “natural” interaction with the instrument.

10.3.3 Visual feedback

Especially in computer music visual feedback is necessary to ensure sufficient transparency in the music creation process. In [Jor01], [JGAK07] the importance of visual feedback to communicate information about instrument state, activity and musical processes. In addition it enables the audience to "watch" music and how it is being constructed. With regard to the musical interaction with a virtual environment the visual representation is of special importance as it provides a visual description of the auditory virtual environment.

Noisternig [NKSS08] describes a real-time auralization approach including a geometric scene graph and 3D visualization. The scene graph provides the description of the geometric room model, including geometry, sound source and receiver positions. The 3D visualization enables an interactive visualization of the room and propagation paths. This concept of providing a scene graph and visualization of the room is considered further for the development of the visual feedback for the instrument and is discussed in chapter 12.5.

10.4 Performer and instrument

As performing music typically is a collective event the development of distributed instruments or providing a collective performance with one instrument is considered in the design process [KJGA06].

The degree of professionalism of the performer is a crucial factor in the interaction design process [Jor01]. An instrument which is e.g. designed for visitors of an interactive sound installation within a public space presumes simply structured interaction processes and comprehensive and expressive instrument output. For a professional musician playing this instrument might not be challenging. Thus the aim in instrument design is to create sophisticated instruments enabling both intuitive interaction and output of complex sound characteristics.

10.5 Playability and apprenticeship of an instrument

In investigations of digital instrument development [Jor04a] focus on requirements of musical instruments in terms of playability and apprenticeship. A basic aim is to find the right balance between challenge, frustration and boredom of the performer in the learning process and thus to derive an appealing and intuitive manageable instrument.

In order to derive this balance the learning curve concept is applied to illustrate the playing effort in time compared to an achievable instrument output efficiency which is used in the context of acoustic musical instruments.

In figure 7 the piano shows a steeper learning curve compared to the violin, which implies rapid learning process within a short period of time. With the piano a high level of efficiency is achievable which indicates high sonic richness and output diversity.

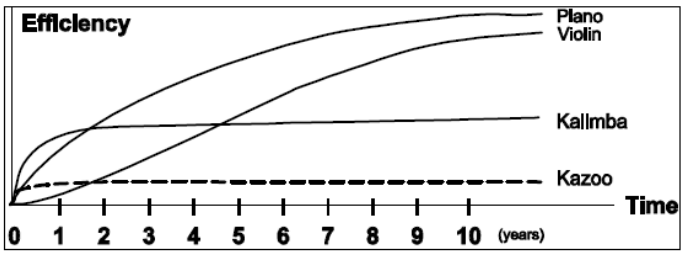


Figure 7: Approximate learning curve for the (a) kazoo, (b) kalimba, (c) piano and (d) violin, within a period of 10 years.[Jor04a]

In order to derive a well shaped learning curve in the context of digital instruments Jordà investigates the instrument efficiency in terms of requirements concerning input control options of the instrument in relation to the derived output complexity.

10.5.1 Musical output complexity

The musical output complexity depends on the range of available sound aspects and their variability properties. In acoustic instruments the sound aspects can be classified in terms of dynamic range, pitch range, resolution of both factors and timbre. In the context of digital instruments the output a general distinction [Jor04b] in terms of diversity, variability and linearity. An instruments with high *stylistic diversity* can be played in various music styles, in contrast low *stylistic diversity* indicates high specialized musical output. An instrument providing high *performance diversity* enables performing the same composition with subtle nuances. The term *variability* indicates the degree of musical expressivity and is connected to the *linearity* in the instrument behavior which enables the exact reproducibility of a musical piece and represents a basic quality criterion in instrument design. A certain non-linear, but predictable instrument behavior can contribute to enrich the sonic pattern of the instrument.

10.5.2 Input control options

The control options of an instrument depend on the available degrees of freedom of control and the number of identifiable gestures and nuances of gestures to control them. An important fact is if simultaneous parameter control and continuous parameter modulation is possible. The quality of the applied mapping strategy is important and should enable the user to perceive the relationship between gesture and audible result in the performance.

10.6 Conclusion of instrumental interface

In the following the advantages of the tabletop tangible interface structure are summarized:

- Controller and sound generation merged into one unit emphasizes the instrument character in the traditional sense.

- Physical representation of digital information implies a high quality of physical interaction with the instrument. In addition the structure enables a topological construction of the instrument during the performance.
- Visual feedback on the table surface communicates information concerning instrument state and current processes.
- Combination of direct-touch and object tracking enables a high number of identifiable gestures.
- Direct-touch bi-manual and multi-finger interaction corresponds to interaction with acoustic musical instrument.
- Setup of an appealing instrument for both, professional and non-professional users, enabling intuitive interaction is provided.

11 Framework of a tabletop tangible user interface (TUI)

In this chapter the components of a tabletop tangible user interface enabling tangible and direct-touch interaction is described. The Reactable is introduced in as example of a practical approach in chapter 11.4.

As mentioned in chapter 10.3 the main parts of the structure are the identification and motion tracking of objects and the transport of the tracking data to a client application. The hardware components of the TUI consist of the table construction, a translucent table surface, the camera which is capturing the table surface and a video projector projecting the graphical feedback onto the table surface. In order to avoid an interference of visual feedback and object/finger detection the two visual tasks need to operate in different spectral bands, which means the camera has to work in the infrared spectrum only.

11.1 Object tracking algorithm

The computer vision algorithm searches each camera image for fiducial symbols and disposes data about fiducial identity, position and orientation. Initially the source image frame is converted to a black and white image using an adaptive threshold algorithm. The image is further segmented into a region adjacency graph which represents the hierarchic structure of interleaved black and white image regions. This graph is searched for the unique tree structures of the fiducial symbols. Matching tree structures are subsequently compared to a repertoire of fiducial symbols to identify the objects.

11.1.1 D-touch

The basement of the computer vision algorithm is the d-touch object tracking algorithm and d-touch fiducial marker system [CSR03] introduced by Costanza and Robinson. In figure 8 a d-touch fiducial symbol and the corresponding adjacency graph in figure 9 are illustrated.

Figure 8: D-touch fiducial marker [CSR03]

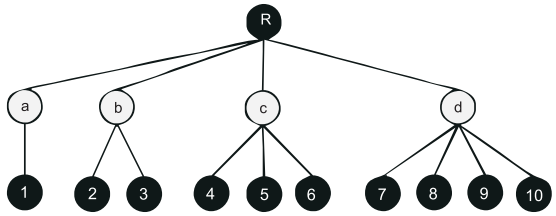


Figure 9: Region adjacency tree for fiducial in figure:8[CSR03]

The d-touch system was employed for the initial implementation of the Reactable, see chapter 11.4. The tracking algorithm was adapted to achieve an increase in processing speed and the marker size and geometry was redesigned as well to match the requirements of real-time musical interaction. The redesigned structure of the fiducial marker results in the amoeba marker set.

11.1.2 Amoeba fiducial marker

The compact symbol structure of the amoeba fiducials was derived using an genetic algorithm [BK05]. A fitness functions was applied to optimize the shape, foodprint size, center point and rotation angle accuracy of the markers. Figure 10 illustrates several simple image structures and their corresponding region adjacency graphs.

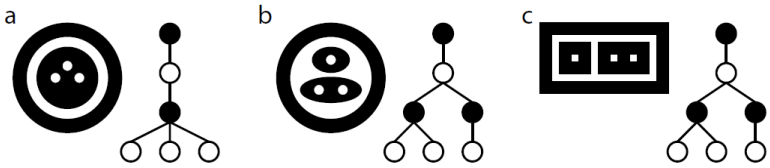


Figure 10: Region adjacency graphs of simple topologies [BK05]

Contrary to the d-touch system the symbols of the amoeba fiducial set are identified only through their topological structure, without using geometric properties for their identification. As shown in figure 11 the position of the amoeba symbol is calculated as the centroid of all found leaf nodes representing the smallest regions in the structure of the symbol. The orientation of the marker represented through a vector pointing from the centroid of the marker to the centroid of a number of black leaf nodes distributed in the upper part of the symbols.

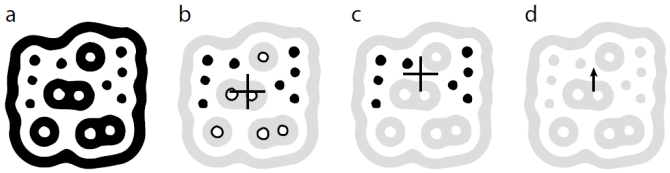


Figure 11: (a) an amoeba fiducial (b) black and white leafs and their average centroid (c) black leafs and their average centroid, and (d) the vector used to compute the orientation of the fiducial [BK05]

There were 128 square fiducial symbols produced for the Reactable application, see chapter 11.4, and 89 symbols were chosen for the implementation depending on requirements concerning size limits and orientation vector length. All symbols have 19 leaf nodes and a maximum tree depth of two. In figure 12 the fiducial with identity no. 1 of the amoeba fiducial set is illustrated.

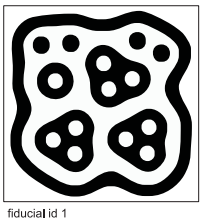


Figure 12: Fiducial with identity no. 1 of “amoeba” fiducial set [BK05]

11.1.3 Multi-finger detection

In order to enable direct-touch multi-finger tracking the basic symbol in the amoeba symbol set represents a white blob corresponding to a single tree branch. This structure implies no calculation of rotation angle, only the finger position, which represents a pointer on the table surface is available. One drawback in including the pointer symbol into the fiducial set is the possible detection of false positives due to noise components in the camera images.

11.2 Communication protocols

The communication protocols used within the TUI structure meet the special requirements of tabletop tangible interfaces and provide fast and reliable communication with local and remote client applications.

11.2.1 Open Sound Control (OSC)

The Open Sound Control protocol was developed 1997 from Wright and Freed at the University of California in Berkeley to support the communication among computers, sound synthesizers and other multimedia devices. A functional overview is given in [WFM03] which is summarized in the following descriptions. A detailed description of OSC 1.0 and OSC 1.1 specification is available opensoundcontrol.org

The OSC data structure represents a client/server structure. The sending application is the *OSC client* and the receiving application represents the *OSC server*. One data unit of the OSC data is called *OSC packet*. OSC is transport-independent which means that the OSC protocol does not specify which underlying network protocol is used to deliver the OSC data from client to server. With a data transmission via Ethernet the OSC packets are sent and received using the User Datagram Protocol (UDP), which is described in chapter 11.2.3. The basic content of a packet is an *OSC message* and consists of address pattern, a type tag string, and arguments. An example of a OSC message address pattern is:

```
/voice/3/freq
```

The control data of an OSC server is organized in an hierarchic tree-structure called address space. The OSC address is the full path from the root of the address space tree to a particular node with a structure format comparable to an URL or file system pathname. The type tag string defines the data type of each argument. The arguments contain the data of the OSC message. Programming environments for the implementation of OSC are e.g Max/MSP, SuperCollider, Open Sound World, Pure Data, Virtual Sound Server, Csound etc. The features of OSC are listed as follows:

- Implementation-defined, dynamic, URL-style symbolic address space.
- Symbolic and high resolution numeric argument data.
- Multiple recipients of a single message can be specified due to the pattern matching language style.
- High resolution time tag to define the absolute time at which the message of a bundle should take effect.
- Transmission of "bundles", which are sequences of messages whose effect most occur simultaneously.
- Query system to dynamically determine the capabilities of an OSC server and to achieve documentation.

11.2.2 TUIO protocol

TUIO [KBBC05] defines a protocol to support the communication between tangible tabletop controller interfaces and underlying application layers. The protocol was initially implemented within the Reactivision framework, see chapter 11.3. TUIO is encoded using the OSC format which is described in chapter 11.2.1. The two main classes of messages are the "set" messages and "alive" messages. The set messages communicate information about the state of an object such as position, orientation and other recognized states. Alive messages indicate the current set of objects present on the surface. There are no explicit "add" or "remove" messages included, thus the receiver deduces object addition and removal by examining set and alive messages. To provide low latency communication UDP transport is used, which implies that several packets may be lost during transport. To correct possible lost packets redundant information is included by the protocol. Each message bundle starts with a "source" message to identify the TUIO source and is concluded with an FSEQ (frame sequence ID) message, which is an incrementing number to provide information about

the consistency of the data flow.

Example bundle structure referring to two-dimensional cursor profile:

```
/tuio/2Dcur source application@address
/tuio/2Dcur alive s_id0 ... s_idN
/tuio/2Dcur set s_id x_pos y_pos x_vel y_vel m_accel
/tuio/2Dcur fseq f_id
```

Profile to define object, cursor and blob descriptors in the case of a 2D interactive surface:

```
/tuio/2Dobj set s i x y a X Y A m r
/tuio/2Dcur set s x y X Y m
/tuio/2Dblb set s x y a w h f X Y A m r
```

s	Session ID (temporary object ID)	int32
i	Class ID (e.g. marker ID)	int32
x, y, z	Position	float32, range 0...1
a, b, c	Angle	float32, range 0..2PI
w, h, d	Dimension	float32, range 0...1
f, v	Area, Volume	float32, range 0...1
X, Y ,Z	Velocity vector (motion speed and direction)	float32
A, B, C	Rotation velocity vector (rotation speed and direction)	float32
m	Motion acceleration	float32
r	Rotation acceleration	float32
p	Free parameter	type defined by OSC message header

Table 2: Semantic types of set messages [KBBC05]

11.2.3 User datagram protocol (UDP)

UDP is a member of the Internet Protocol Suite and uses the Internet Protocol (IP) as underlying protocol. It uses a simple transmission model to send datagrams to host applications without guaranteeing reliability, ordering or data integrity. In real-time applications UDP is used because it is preferred to drop data instead of waiting for delayed data.

11.3 Reactivision framework

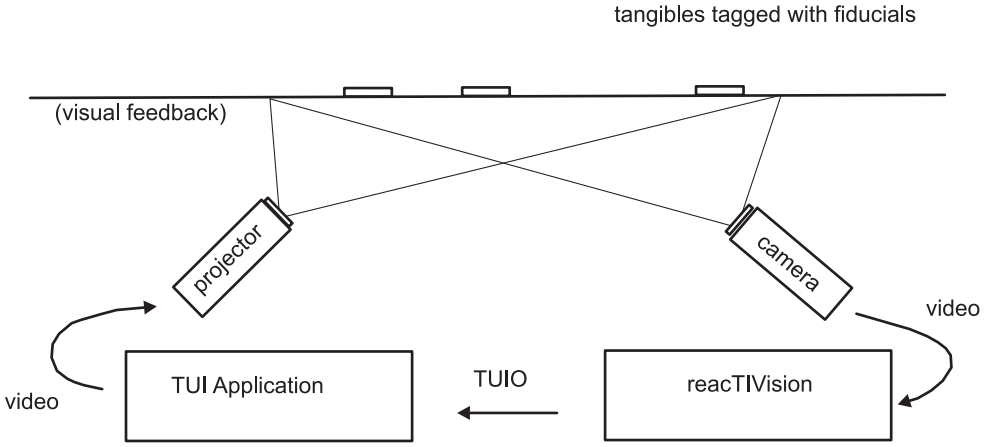


Figure 13: Reactivision diagram [KB07]

The Reactivision framework [KB07] is an application framework to support the construction of table based tangible user interfaces. It has primary been developed for the Reactable, which is described in the following chapter. The framework provides a basic open-source software structure and combines the components described in chapter 11.1 and 11.2. It is written in C++ with builds for all three major operating systems Linux, Mac OS X and Windows. The video acquisition is based on PortVideo which is also available open source.

11.4 Example approach - Reactable

Several examples of tangible musical interfaces are the Audiopad [PRI02], Xenakis [BCL+08] and the Reactable [JKGB05] which is probably the most famous implementation and described below. There is an overview given by Kaltenbrunner [Kal10].

The Reactable is a multi-user electro-acoustic music instrument based on a tabletop tangible user interface. It was developed by Sergi Jord, Martin Kaltenbrunner, Gnter Geiger and Marcos Alonso at the Pompeu Fabra University in Barcelona since 2003. The tangible artifacts on the table surface are physical representations of the components of a classic modular synthesizer. The tangible interaction directly controls the topological structure and the parameters of the synthesizer. Thus the performer constructs and plays the instrument at the same time. In order to obtain a collaborative performance several simultaneous performers share the control over the instrument. The visual representation communicates information about current instrument state, activity and visualize the main characteristics of the sounds produced by the synthesizer.

Part III
Practical contributions

12 Instrument setup and playing paradigm

In this chapter the development and playing paradigm of the instrument is described. Three functional units are implemented: (1) hardware setup of the instrumental interface enabling tabletop tangible and direct-touch interaction, (2) development of the tangible object functionality including the development of a parameter mapping strategy, the derivation of musical control parameters and the setup of a room acoustic model enabling real-time auralization of the AVE, and (3) the visualization of the interaction process on the table surface.

12.1 Setup of instrumental interface

The instrument is based on the structure of a tabletop tangible interface described in chapter 10.3.1. The implementation of both, tangible interaction and multi-touch functionality, which enables the detection of multiple fingers on the table surface, provides a considerable number of different identifiable performing gestures. Figure 14 illustrates the camera image of the interactive surface. The green symbols indicate the recognition of the amoeba fiducial identity no. five and the detection of four fingers. Figure 15 illustrates the black/white camera image, after the conversion by using the adaptive threshold filter.

Figure 14: Recognition of fiducial symbol and finger events

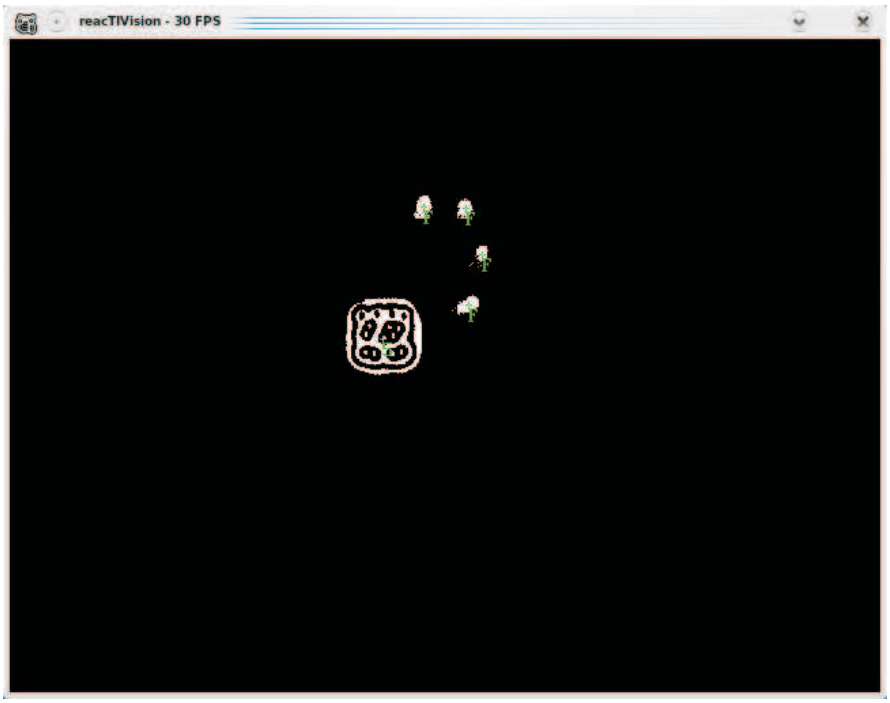


Figure 15: Grey scale image converted into black white image

A detailed description of the acquisition of the tracking data is provided in the software implementation of the interface in chapter 13.2. The surface of the interface provides a projection screen for the visualization of current instrumental processes in order to support the interaction process between performer and instrument. The detailed description of the hardware setup of the interface is provided in chapter 14.

12.2 Tangible object functionality

In the performance of the instrument the auditory virtual environment is composed by the arrangement of tangible objects on the surface of the interface. The tangible objects imply the functionality of virtual sound sources and virtual architectural elements and embody the control units for the configuration of the virtual space. In order to define the functionality of the objects at first their position data is mapped to control parameters and the functionality of the particular tangible objects is defined.

12.2.1 Development of a gesture mapping strategy

In order to enable the playing of the virtual room as musical instrument adequate musical control parameters have to be derived. Thus a mapping strategy between gestural parameters and control parameter needs to be developed. Each fiducial object provides five gestures which can be converted into instrumental control parameters. In table 3 the different gestures and their parameter representations are listed. Figure 16 illustrates the dynamic data derivation in case of a fiducial object.

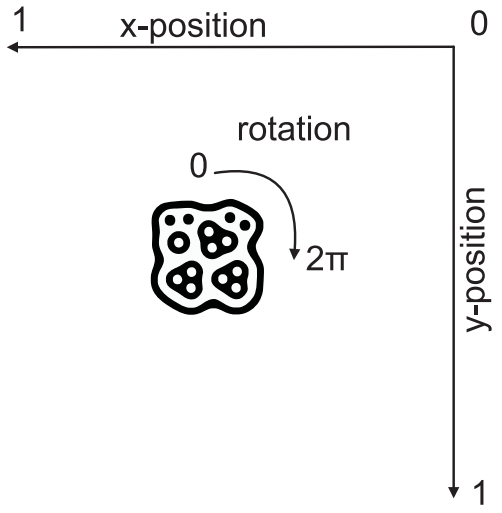


Figure 16: 2D position and rotation data of fiducial object

Adding an object on the table surface	/addObject/object-identity
Removing an object from the table surface	/removeObject/object-identity
Dynamic x-coordinate of fiducial position	/updateObject/object-identity/x-position
Dynamic y-coordinate of fiducial position	/updateObject/object-identity/y-position
Rotation value of fiducial object	/updateObject/object-id/rotation
Finger detection	/addCursor/finger-identity
Remove finger	/removeCursor/finger-identity
Dynamic x-coordinate of finger	/addCursor/x-position
Dynamic y-coordinate of finger	/addCursor/y-position

Table 3: Identifiable gestures and according parameter description

These parameters are employed to develop the functionality of virtual sound sources and virtual room geometries. Thus the tangible objects are defined to represent either sound source or wall objects. The particular functionality is assigned by the object-identity, which is the identity of an fiducial object as explained in chapter 11.1.

12.2.2 Sound source object parameters

The "/addObject" message in combination with an object-identity, assigned to the functionality of a source object, activates the sound source by starting a sound file player. The "/updateObject" message of the particular object contains the x- and y-position data and the rotation value of the object. The 2D position data is converted into spherical coordinates, considering a 2D projection of a hemisphere, which is described in chapter 13.2.2. The position data defines the position of the virtual sound source in azimuth and elevation for the Ambisonics encoding. The rotation angle is assigned to the control of the sound level of the source signal and delivers the gain parameter of the direct sound source signal. With the removal of the source object the "/removeObject" message stops the sound file player. By exploiting the

advantage of tangible interaction the control of multiple simultaneous sound sources is enabled. The following list shows the derived parameter of each sound source object.

Definition of Virtual Acoustic Instrument (VAI) data protocol:

Action on surface	Control parameter	Parameter space
AddObject	Activation of auralization	/EXT/VAI/SRC/ID/ON
x/y-pos (spherical)	Azimuth angle	/EXT/VAI/SRC/ID/AZM
x/y-pos (spherical)	Elevation angle	/EXT/VAI/SRC/ID/ELV
Rotation angle	Source volume	/EXT/VAI/SRC/ID/GAIN
RemoveObject	Deactivation	/EXT/VAI/SRC/ID/OFF

Table 4: Virtual Acoustic Instrument (VAI) parameter space, sound source

The parameter name space is chosen in order to enable the connection of the software implementation, to the Cubemixer software, which is explained in chapter 13.1.6.

12.2.3 Wall object parameters

Tangible architectural objects enable a dynamic modification of the virtual room geometry and absorption properties. Each wall object represents a particular wall of the virtual environment. With the "/addObject" message in combination with the particular wall object-identity a particular wall is activated. The rotation value is assigned to the modification of the virtual distance from the wall to the performer. The following parameter are defined by the functionality of a wall object:

Action on surface	Control parameter	Parameter space
AddObject	Activation virtual wall	/EXT/VAI/WALL/ID/ON
Rotation angle	Distance wall-receiver	/EXT/VAI/WALL/ID/DIST
RemoveObject	Deactivation	/EXT/VAI/WALL/ID/OFF

Table 5: Virtual Acoustic Instrument (VAI) parameter space, wall element

12.2.4 Direct-touch finger parameters

The direct-touch functionality is employed to enlarge the number of recognizable gestures on the interactive surface. In the implementation of the instrument it is employed to augment the functionality of a tangible wall object. Thus in the combination with the activation of a wall object the selection of three different wall preset options defining the absorption property of the virtual wall is enabled, as illustrated in figure 18. The detection of a finger inside a defined area releases the setting of a particular preset and defines the parameters listed in the below table. Definition of Virtual Acoustic Instrument (VAI) data protocol:

Action on surface	Control parameter	Parameter space
Finger Position	Absorption coeff. 1	/EXT/VAI/WALL/ID/PRESET 1
Finger Position	Absorption coeff. 2	/EXT/VAI/WALL/ID/PRESET 2
Finger Position	Absorption coeff. 3	/EXT/VAI/WALL/ID/PRESET 3

Table 6: Virtual Acoustic Instrument (VAI) parameter space, direct-touch

12.3 Room acoustic model - setup and control

The derived parameters indicated in the previous chapters are now employed as musical control parameters for the implementation of the room acoustic model. The sound source in combination with a wall object generates a single reflections and thus auralize the acoustic behavior of a particular wall element. The overall configuration of the wall elements controls the reverberation within the virtual environment. The room acoustic model consists of an image source model computing early reflections and of a feedback delay network (FDN), described in chapter 8.4.2, which simulates the reverberation within the virtual environment.

12.3.1 Modeling of early reflections

With regard to the performance of the virtual architecture as musical instrument first order reflections are generated to depict the acoustic property of single walls. Due to the possibility of configuring the distance and absorption property of each wall the usage of a wall element as musical parameter is provided. With each wall object three different absorption properties can be adjusted. The computation of early reflections is based on a image source model described in chapter 8.2.3 and performs the calculation of first order specular reflection paths. The reflection paths are calculated relative to the static listening position located at the place of the instrument in the center of the performance space. The image model receives the control parameter of the sound source and wall objects. The source position of each source in spherical coordinates are converted into room coordinates according to the descriptions in chapter 13.2.2. Based on the position data the image source positions with regard to modifiable room geometries and their distances are calculated. The distance of each image source is converted into the time delay and attenuation value. The selected wall preset determines the frequency dependent absorption property of each particular wall.

12.3.2 Modeling of late reverberation

The room geometry, resulting room volume and the overall absorption property of the room, defined by the configuration of the wall elements, provides a further musical parameter which enables the control of the reverberation within the enclosed space. Based on the room volume and absorption properties the reverberation time by Sabine, introduced in chapter 4.3.3, is calculated for three frequencies. The derived decay times further provide parameters for the control of the feedback delay network (FDN), described in chapter 8.4.2. Thus an adaptive adjustment of the decay time of the reverberation in three frequency bands is enabled.

12.3.3 Sound spatialization using Higher Order Ambisonics (HOA)

The encoding of the room impulse response is performed by the implementation of a Higher Order Ambisonics (HOA) sound field reproduction, which is introduced in chapter 9.2. Using HOA in auralization systems enables a decoupled encoding and decoding of the sound field, which means the reproduction of the sound field is seperated from the encoding module, which enables the encoding of the sound field independent from the loudspeaker setup in the decoding of the sound field. An additional advantage in HOA is that the spatial resolution of the sound field is a

function of the Ambisonics order. Depending on the required reproduction accuracy a mixed order Ambisonics implementation is possible.

12.3.4 Encoding of the sound field

In order to obtain high localization accuracy in the representation of the direct sound and early part of the room impulse response, higher order encoding is applied. In order to save processing power, the encoding of the diffuse sound field is performed using a lower Ambisonics order.

12.3.5 Decoding of the sound field

The optimal reproduction of the sound field depends on the order applied in the decoder unit. In the implementation of the auditory virtual environment both loudspeaker and headphone reproduction is possible. The derivation of a binaural representation of the sound field, according to the description in chapter 9.6 is derived by filtering the Ambisonics decoded loudspeaker signals with head-related transfer functions (HRTFs), which are described in chapter 5.1.

12.4 Excitation signal

The basic purpose of the excitation signal is to depict the acoustic quality of the virtual environment. Basically there are no limitations of what kind the signals can be: music, speech, synthesized sound, etc. In general signals are of different quality in the evocation of spatial auditory perceptions and should be well motivated to excite the attributes of interest [BZ06, chap. 5.1].

The excitation signal can be classified in terms of the following criteria:

- Signal generation
- Time domain characteristics
- Spectral characteristics
- Direction and movement

12.4.1 Signal generation

Depending on the signal generation the sound signals can be divided into direct sound signals present in the performance room, synthetic signals and recorded signals.

Direct sound signals occuring in the performance space can be singing voice, speaking voice, sound of instruments and noise induced by the performer, soloists and audience.

Synthesized signals are any sound signals produced using analog or digital electronic equipment. Generating a synthetic excitation signal enables the specification of the signal conform to the requirements in auralization of room acoustics. In addition signal synthesis enables the creation of signals which cannot be realized using a direct or recorded sound signal.

Recorded signals need to be recorded under anechoic conditions in order to avoid the superposition of recorded and artificial generated reverberation. Recording technique, storage and reproduction further influence the quality of the sound signal.

Ambient sounds are environmental sounds and are employed in order to create an auditory impression associated with a certain environment. The sounds can either be synthesized or recorded and are e.g. sound of birds, traffic, the sound of waves, etc.

12.4.2 Time domain characteristics

The structure of the signal as function of time can be described in terms of the degree of stationarity and has an influence on the audibility of various aspects of the sound signal. With regard to room acoustics stationary signals and time varying (impulse) signals are of different quality in the auralization of resonances.

12.4.3 Spectral characteristics

The spectral component of a sound signal is important, as mentioned in the theory of spatial hearing, see chapter 5, for the localization of the sound source in the median plane. In addition the excitation of room modes depends on the frequency of the excitation signal. As explained in chapter 8.4.1 the density of modes (eigenfrequencies) increases with frequency, independent of the room shape.

12.4.4 Static and dynamic sound sources

In real environments sound is occuring from both, static and dynamic sound sources in the horizontal and the median plane. In terms of creating a convincing auditory environment and in order to achieve an appealing instrumental output, the spatial rendering introduced in chapter 9.2 is applied. It enables the positioning and movement of the sound sources within the horizontal plane with an azimuth angle φ from 0 to 360 and in the median plane with an elevation angle δ from 0 to 90.

12.5 Development of visual feedback

The development of a comprehensive and representational visual feedback is of basic importance in order to support the interaction process between performer and instrument. According to the descriptions in chapter 10.3.3 Noisternig [NKSS08] proposes a real-time aurzalization framework including a scene graph unit based on Virtual Choreographer [Cho06]. In this approach the data for the room acoustic model is provided by the scene graph unit, which delivers data about the room geometry, sound source(s) and receiver(s) positions. It further provides the visualization of the room and the propagation paths derived by the room acoustic model. In the development of the visual feedback of the instrument this concept is adapted. Thus a visual representation of the virtual scene and a visual information about the dynamic modification of the scene is provided on the surface of the interface. Two forms of visual feedback, the static visual representation of the scene graph of the virtual environment and a temporary, dynamic visual representation according to the activity on the interactive surface is provided.

12.5.1 Static representation

The static elements of the scene graph illustrate a 2D view of the modifiable geometry of the virtual environment and define the active area in which the sound sources can be activated. Level meter are displayed to provide information about the gain of the sound sources. Figure 17 shows the visualization of the geometric scene graph and level meter.

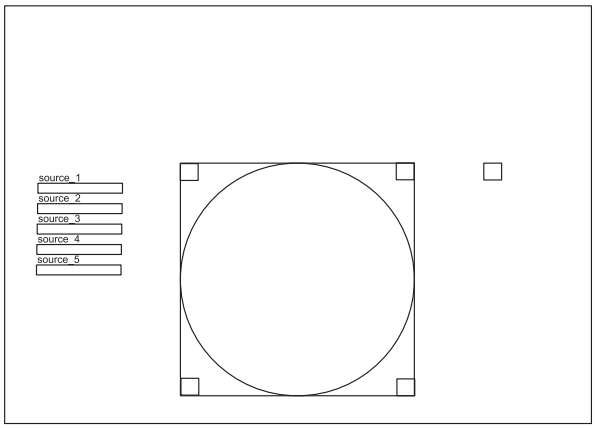


Figure 17: Static visualization of 2D scene graph

12.5.2 Temporary, dynamic visual representation

In order to provide visual information about the actual state of the instrument each sound source and wall object disposes of a temporary, dynamic visual representation. Thus an additional parameter mapping of the, in chapter 12.2.1 described object parameters, to derive the visual feedback is developed.

In the case of a sound source the "/addObject" message activates a temporary visualization of a filled circle surrounding the object which confirms the recognition of the source object. The position data of the "/updateObject" message is converted according to the description in chapter 13.2.2 and controls the motion of the visualization conform to the movement of the tangible source object. The rotation value indicates the sound level of the source signal and controls the display of the level meters.

With the "/addObject" message of a wall object the activation of the particular wall is highlighted. The actual distance of the wall to the performer is displayed by the indication of the distance value in meter. In addition the visualization of the finger-touch sensitive areas which enable the selection of the preset options are provided. Figure 18 and figure 19 illustrate the activated dynamic visualization.

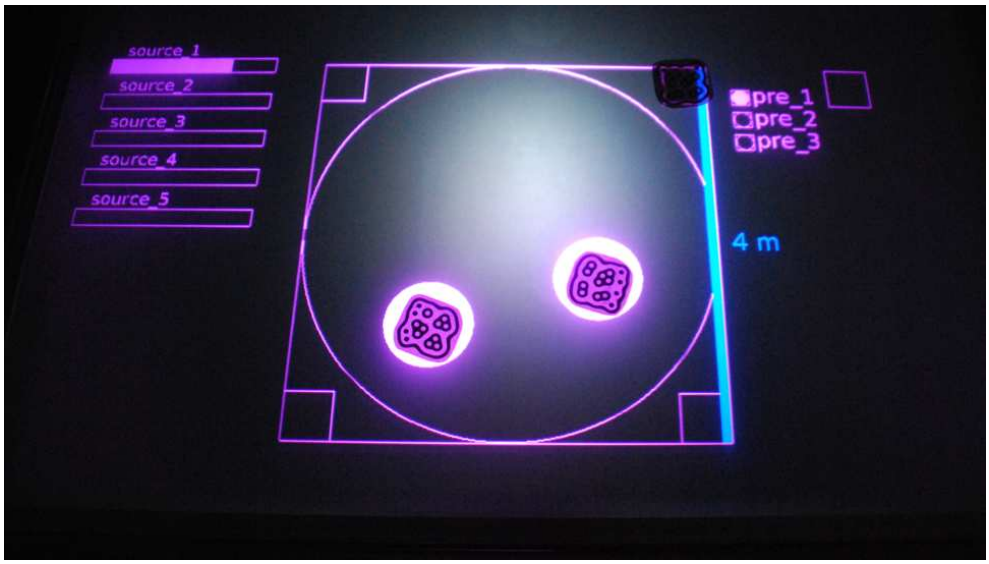


Figure 18: Visualization of source and wall objects

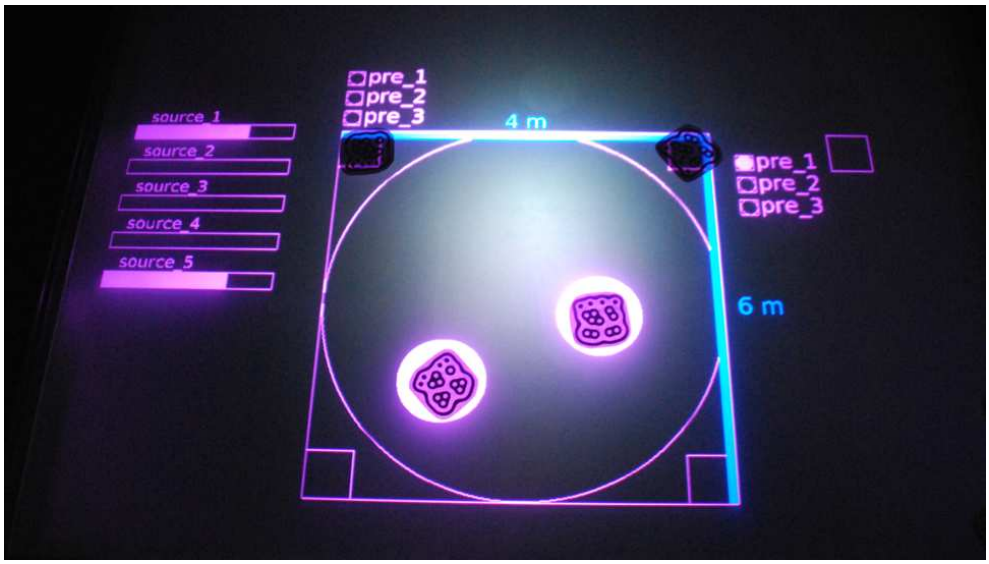


Figure 19: Visualization of source and wall objects

13 Software implementation of the instrument

For the software implementation of the instrument two software modules were developed. The first module implements the functionality of the TUI which handles the acquisition of the tracking data and the generation of the visual feedback. The second module receives the tracking data of the tangible user interface (TUI) and converts the data into the control parameter of the instrument which further control the room acoustic modeling and reproduction of the sound field.

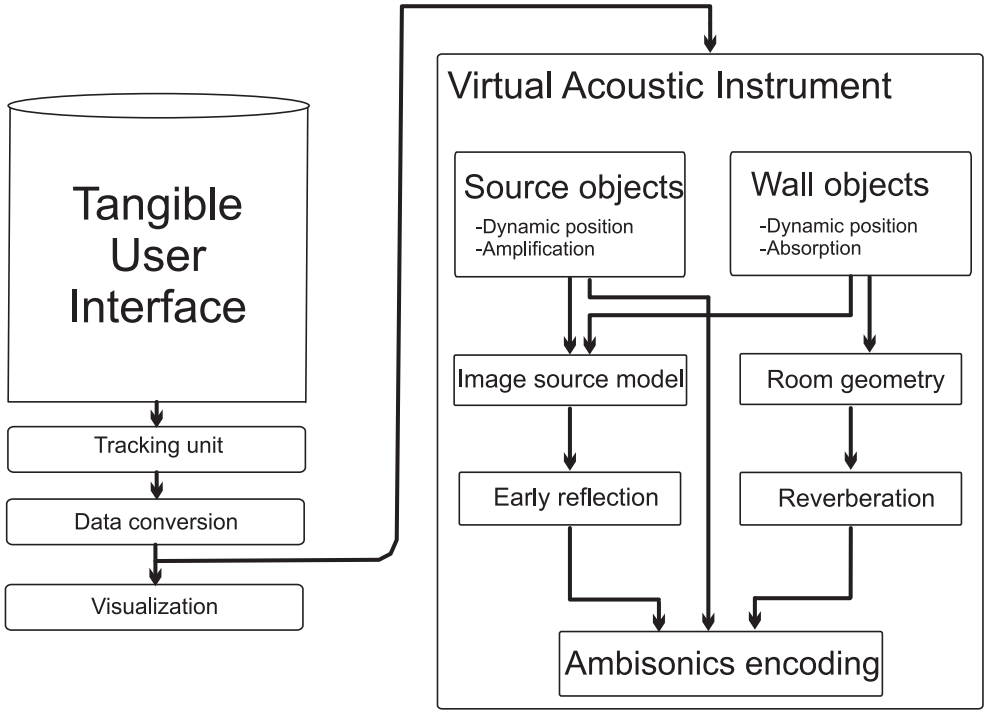


Figure 20: Instrument software components

13.1 Software deployment

The open-source operating system Linux with the Ubuntu studio, 64 Bit distribution, with KDE desktop environment is employed. Ubuntu studio provides installed open-source applications for audio, video and graphics processing. For software implementation of the instrument is based on Pure Data (PD), introduced in the following chapter using the latest version (Feb. 2010) 0.41.4 for Linux. The graphics processing is performed by the Graphics environment for multimedia (Gem), which is the graphics library of PD, see chapter 13.1.2. The software implementation representing a collection of PD extensions, is available open source and can be downloaded via subversion checkout from the repository: https://svn.iem.at/svnroot/iem/iemtable/.

13.1.1 Pure data (PD)

PD is a real-time graphical programming environment developed by M. Puckette et al. at the IRCAM [Puc96]. It enables audio, video and graphical processing in real-time and is a member of the family of patcher programming languages known as Max. PD is a free software and available on the operating systems GNU/Linux, Windows, IRIX, BSD and Mac OS X. Process scheduling in PD is divided in two layers, the signal layer and control layer. The siganl layer enables data processing in real-time synchronous to the signal input and output device of audio or graphics card. In the control layer message data is processed asynchronous on demand. This enables a minimization of the computing time.

13.1.2 Graphics environment for multimedia (Gem)

Gem is a collection of PD externals to provide OpenGL based three-dimensional graphics processing in PD and was originally developed by Mark Danks. The externals support polygon graphics, lighting, texture mapping, image processing and camera motion and the application follows the programming paradigm of PD.

Open GL is a standard specification defining a cross-language and cross-platform application programming interface (API) for writing applications which produce 2D and 3D computer graphics. The basic operation is the conversion of geometric primitives, such as points, lines and polygons into pixels. The conversion is accomplished by a graphics pipeline known as OpenGL state machine. Over 250 function calls enable the drawing of complex three-dimensional scenes from simple primitives. The two main purposes of OpenGL are to provide a single, uniform API in order to avoid interfacing with different 3D accelerators and to cope with different capabilities of hardware platforms.

13.1.3 Jack Audio Connection Kit

JACK [Dav06] is a system for handling real-time, low latency audio (and MIDI). It is available for the operating systems GNU/Linux, Solaris, FreeBSD, OS X and Windows. It can connect a number of different applications to an audio device, as well as allowing them to share audio between themselves. Its clients can run in their own processes (e.g. as normal applications), or can they can run within the JACK server (e.g. as a "plugin"). JACK supports distributing audio processing across a network, for both fast and reliable local area networks as well as slower, less reliable wide area networks.

13.1.4 Communication between software modules

The communication between the software modules is enabled using Open Sound Control (OSC) [WFM03], explained in chapter 11.2.1, and the UDP protocol, in chapter 11.2.3. The usage of OSC allows the distribution of the software modules to separate processors which are connected via local area network or Ethernet. Thus a higher processing performance and the possibility of remote controlling of the instrument is provided.

13.1.5 Latency

The latency of the instrument depends on the following software parts: latency of the visual recognition part, the operating system, JACK audio connection, Pure Data processing and the performance of the data transfer via Ethernet or local area network. The latency of the visual recognition part depends on the frame rate of the firewire camera, which is in the implementation 30 fps, see chapter 14.2, and results in 33 ms latency. The latency time in Pure Data is set to 40 ms in order to avoid getting an interruption in the audio output. In the audio connection kit the latency is set to 40 ms, without using real-time priority. The Ethernet connection implies an average latency of 2 ms. As mentioned in chapter 10.2 especially in playing a musical instrument, latency is a crucial factor and thus should be kept as low as possible.

13.1.6 Cubemixer software

The software architecture of the second module accomplishing the room acoustic modeling and sound spatialization is based on the Cubemixer software [MRZry] which is mixing and mastering tool for multichannel speaker systems or binaural rendering. The software is based on the Amisonics library for PD and is an open source software which provides a decoupled encoding and decoding unit for the reproduction of the sound field. The Cubemixer software module is further splitted into the audio engine, processing the audio signal and the control engine, processing the control data. The Cubemixer software provides the connection of extension applications. Thus the parameter name space of the software implementation of the instrument is chosen in order to connect the instrument to the Cubemixer software.

13.2 Software implementation of tangible user interface (TUI)

The software module of the TUI provides three main functionalities: (1) The acquisition of the position data of the fiducial object and fingers on the table surface. (2) The conversion of the position data into coordinate systems which further enable the processing of the visualization, room acoustic modeling and sound spatialization. (3) The graphics processing of the visual feedback on the surface of the user interface.

13.2.1 Acquisition of tracking data

The position data of fingers and fiducial objects on the table surface of the TUI are obtained by the Reactivision framework which is described in chapter 11.3. The framework provides the PD external tuioClient.pd, which receives the TUIO messages, described in chapter 11.2.2 and supports the setup of a 2D interactive surface. The derived gesutral paramters are described in chapter 12.2.1 and are shown in table 3.

13.2.2 Data conversion

The position data mapping requires the conversion of the 2D position data, illustrated in figure 16 in chapter 12.2.1, of objects and fingers in order to perform three different purposes:

- Graphics processing PD Gem: Transformation of 2D fiducial coordinates into 3D coordinate system, with origin in the center of the gem window.

- Room acoustic modeling: Transformation into 3D room coordinates.
- Sound spatialization: Transformation of sound source and image source positions into a head-related spherical coordinate system.

The transformation between the spherical coordinate system and the Cartesian coordinate system is performed by using the Pure Data library "zexy". The relation of Cartesian and spherical coordinates is given by [Bar01, p. 214]:

- $x = r \sin \delta \cos \varphi$
- $y = r \sin \delta \sin \varphi$
- $z = r \cos \delta$

where r denotes the radius from the sound source to the receiver position in the origin of the coordinate system, φ corresponds to the azimuth angle and δ represents the elevation, according to the head-related coordinate system introduced in chapter 5, figure 3.

The transformation of the Cartesian coordinates into spherical coordinates is based on a 2D projection of the spherical coordinate system. Figure 21 illustrates a 2D projection of a hemisphere. The azimuth angle is orientated in mathematical negative rotating direction. The elevation angle is geodetical orientated, with 0 elevation corresponding to a positioning at the "equator" and 90 corresponding to a position at the "north pole".

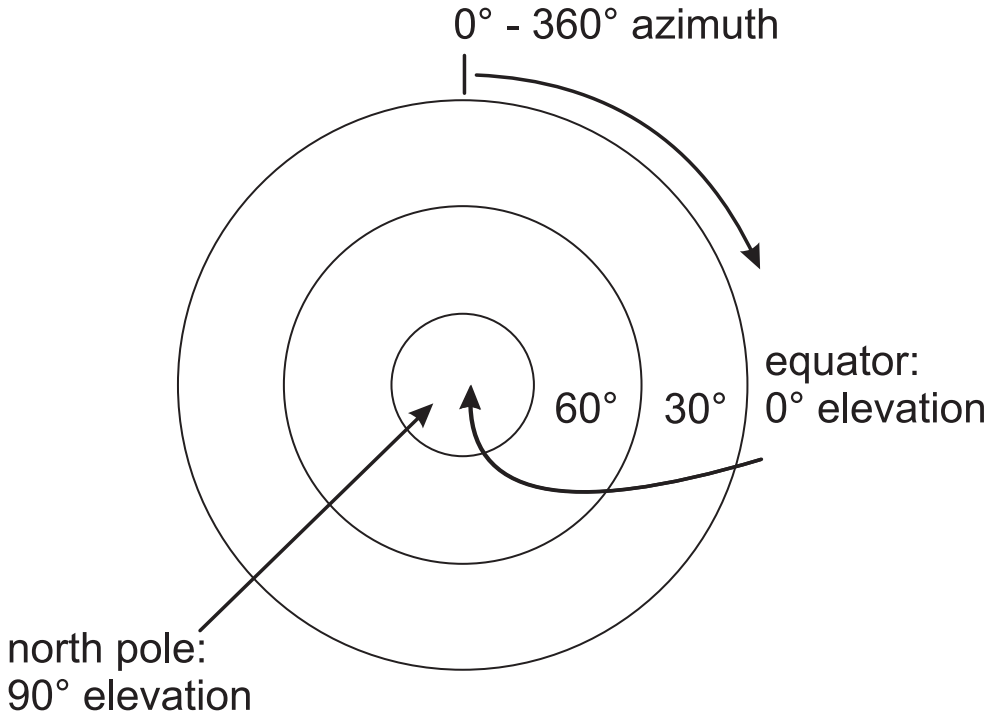


Figure 21: Graphical 2D projection of hemisphere

13.2.3 Graphics processing

The graphics processing is performed by the graphics environment for multimedia (Gem) introduced in chapter 13.1.2. The PD Gem extension gemwin.pd controls the window manager and passes various messages to the manager, controlling the attributes of the window. With the create message the graphics window is created in contrast the message destroy deletes the window. The extension gemhead.pd connects the Gem objects, e.g. circle, rectangle, etc., to the window manager. With receiving a "bang" message the gemhead receives the render command and is displayed in the graphics window. In the implementation of the instrument the graphics window is projected onto the table surface and thus provides the visual feedback on the table surface. In the implementation of the visual representation the basic geometric Gem objects rectangle.pd and circle.pd are applied. As explained in chapter 10.3.3 the visual feedback comprises static and dynamic components. The static Gem objects receive the render command with the initialization of the instrument. In contrast the temporary visualization is activated and deactivated by the "/addObject" and respectively "/removeObject" parameters, explained in chapter 12.5.2. The dynamic movement of the visual objects is enabled by using the translate.pd object. The object accepts a (x,y,z) vector, which is represented by the consistently updated movement data of the tangible objects on the table surface.

13.3 Implementation of the room acoustic model

According to descriptions in chapter 13.1.6 the basic software structure of the room acoustic model is divided into the audio engine and the control engine. The audio engine accomplishes the processing of the audio signals. The control engine of the instrument receives the position data of the tangible objects and computes the control parameter of the instrument which further perform the acoustic modeling of the virtual environment.

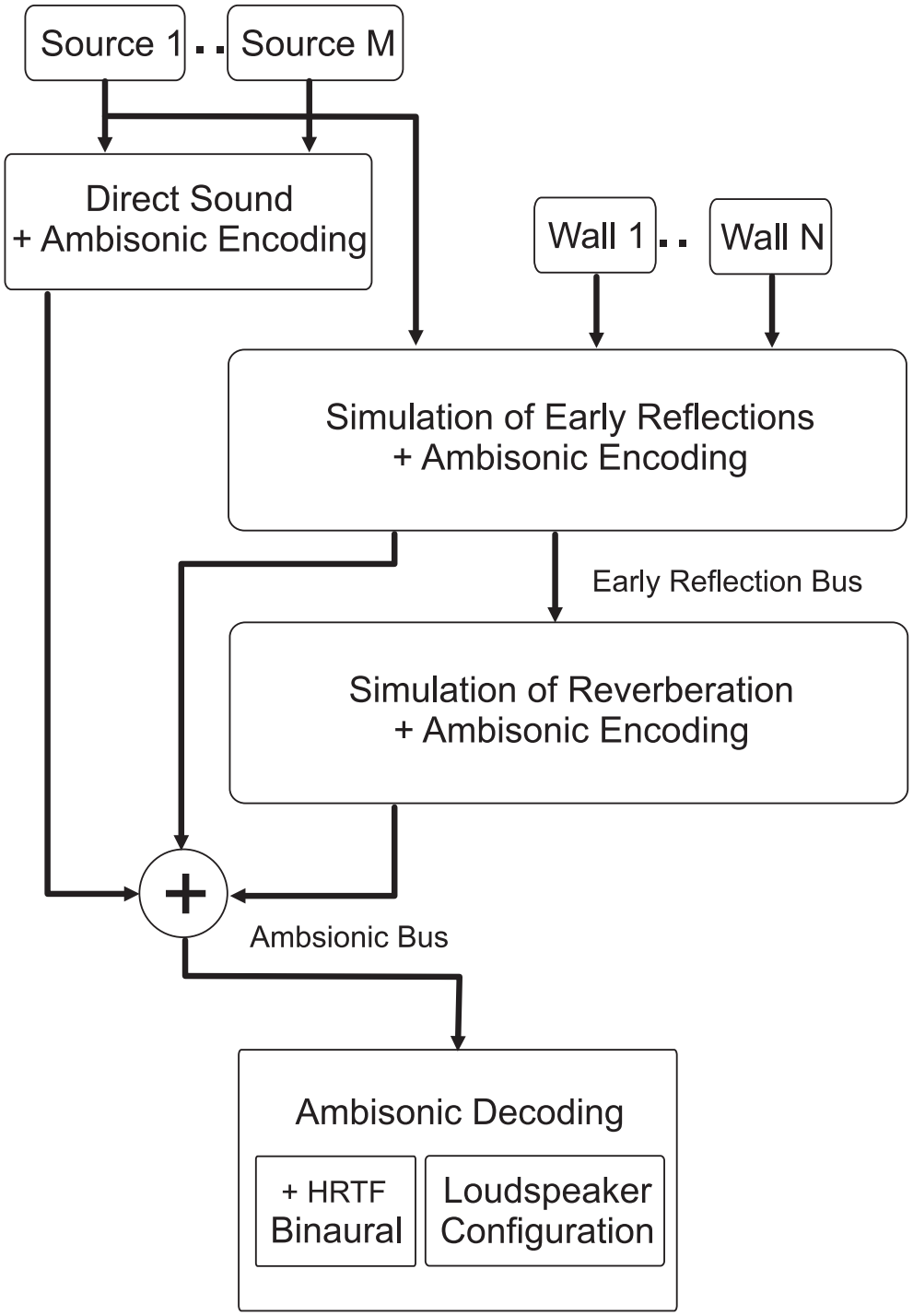


Figure 22: Room acoustic model

The room acoustic model processes the direct sound of M sound sources and

computes the first order specular reflections in a rectangular room geometry with N = 6 walls. The model comprises the direct sound, early reflections and reverberation processing units which dispose of distinct encoding stages in order to enable the scaling of the spatial resolution of the sound field, as explained in chapter 12.3.4.

13.3.1 Early reflection simulation

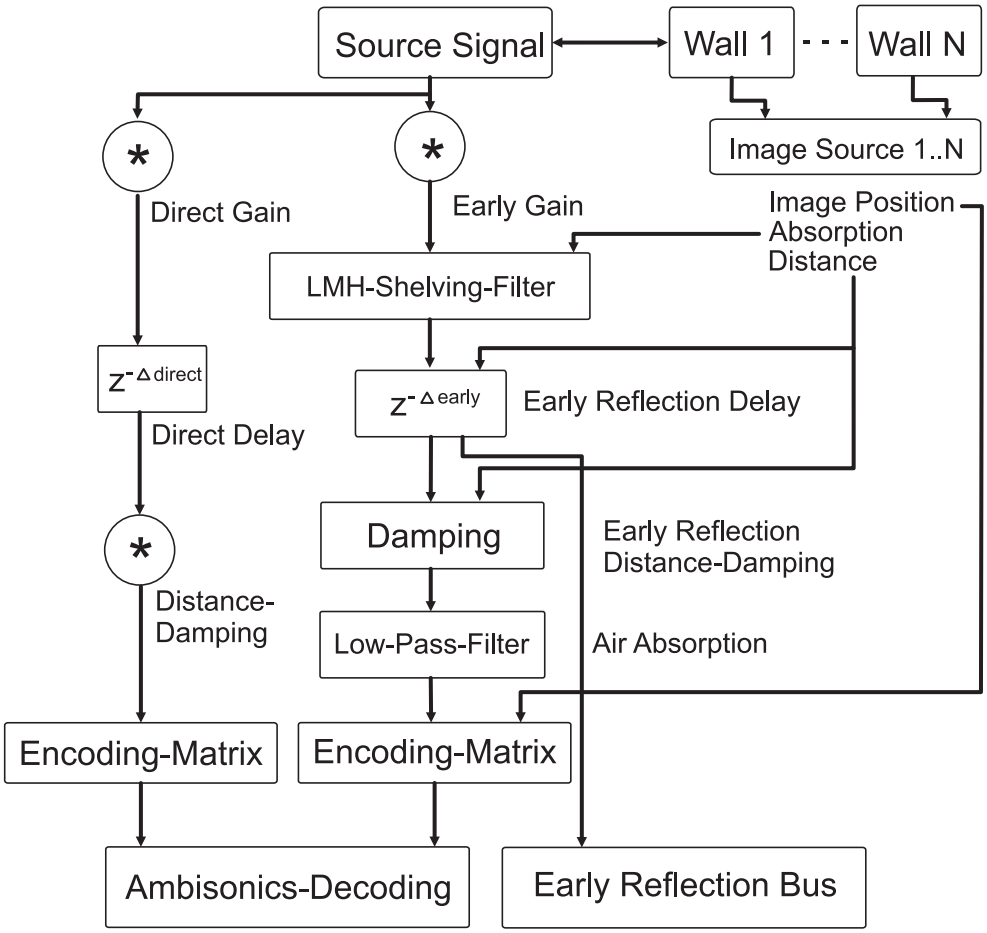


Figure 23: Early reflection simulation

The control parameter of the room acoustic model are derived by the tangible objects on the table surface as explained in chapter 12.2.1. In the binaural case the direct sound source signal is delayed and attenuated according to the distance of the sound source to the receiver. In case of loudspeaker reproduction the minimum radius to the sound source is determined by the loudspeaker positions. The frequency dependent attenuation of the reflection is derived by filtering of the signal with a IIR shelving filter in three frequency broad bands (<500 Hz, 500 to 2000 Hz and >2000 Hz). The air absorption is considered by low pass filtering of the reflections. The length of the reflection path determines the attenuation and the delay of the reflection, which is

calculated by Δt=l/c, where c is the speed of sound and l represents the length of the reflection path. The distance damping of the sound pressure level is given by 1/r. The position of the image source is calculated as proposed by Allen and Berkley [AB79].

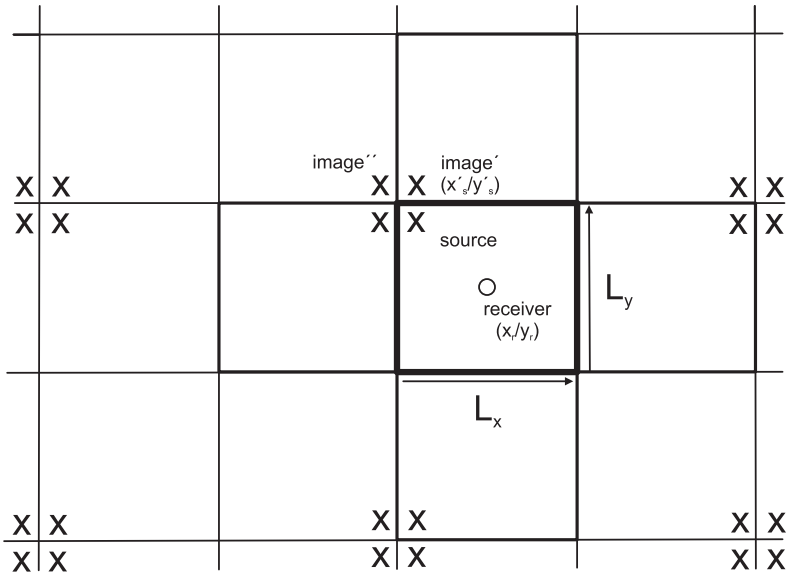


Figure 24: Image source model

Initially the position of the sound source and receiver position is converted to Cartesian room coordinates. The position of the sound source is then mirrored at the x-, y- and z-axis and at the displaced axis considering the distance to the opposite room boarders by the room dimensions L_x, L_y, L_z.

13.3.2 Simulation of reverberation

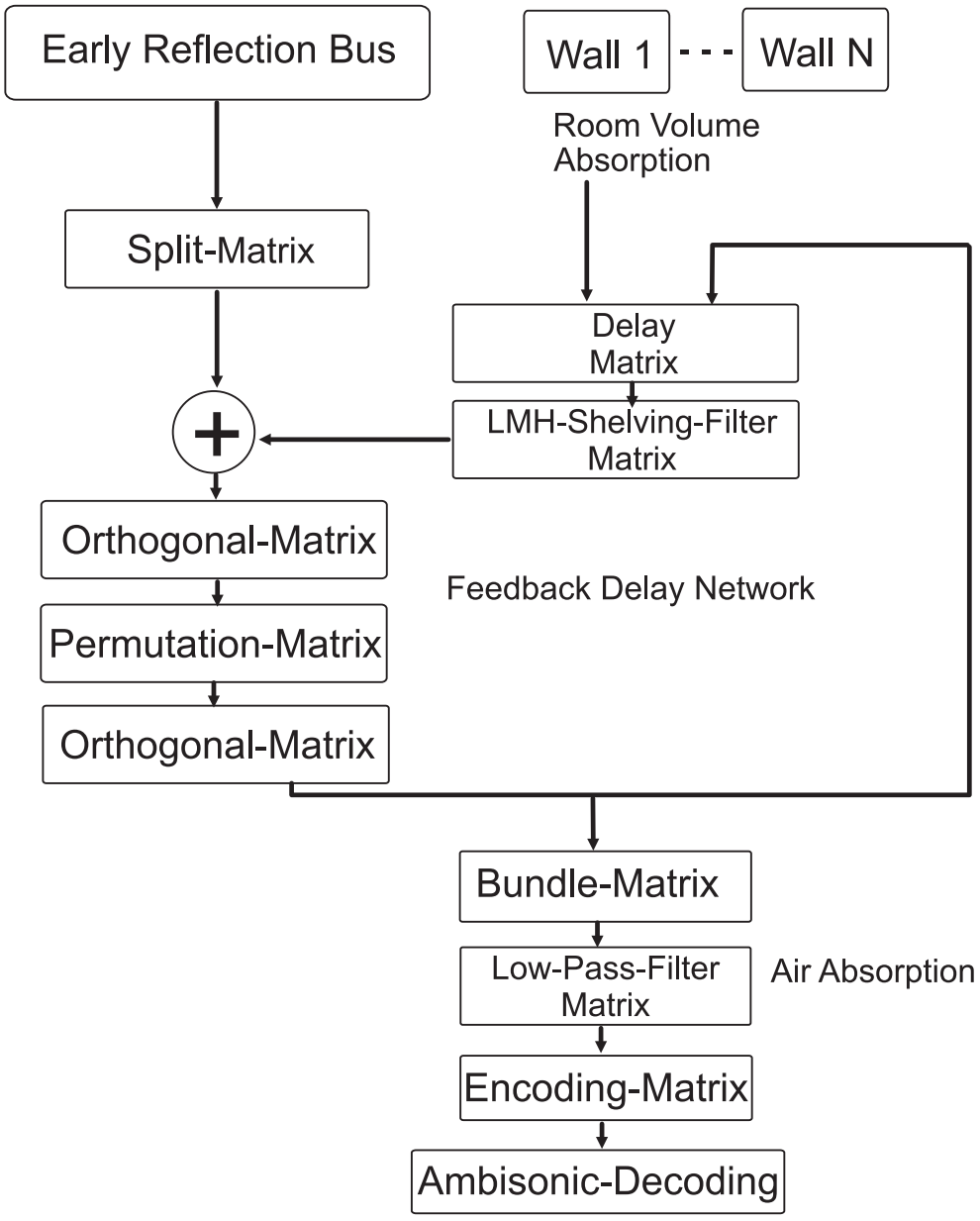


Figure 25: Simulation of reverberation

The implementation of the reverberation is based on the Pure Data library for real-time binaural 3D sound reproduction [MNH05]. The signals of the early reflection bus are spread using a split matrix which provides an initial diffusion of the input signals. The orthogonal matrix in the feedback loop is based on the scheme of a Householder matrix, which is explained in chapter 8.4.2. The absorption property of the room is

simulated by the filtering of the signal with an IIR shelving filter in three frequency broad bands (<500 Hz, 500 to 2000 Hz and >2000 Hz). Additional damping through the wave propagation path is considered by a second order low-pass filter.

14 Hardware setup of the table interface

The hardware setup of the tangible user interface, as described in chapter 11 is based on a table construction, a translucent table surface, the camera which is capturing the table surface and a video projector providing the graphical feedback on the table surface. The system setup is illustrated in figure 26 and the particular components are described in the following chapters.

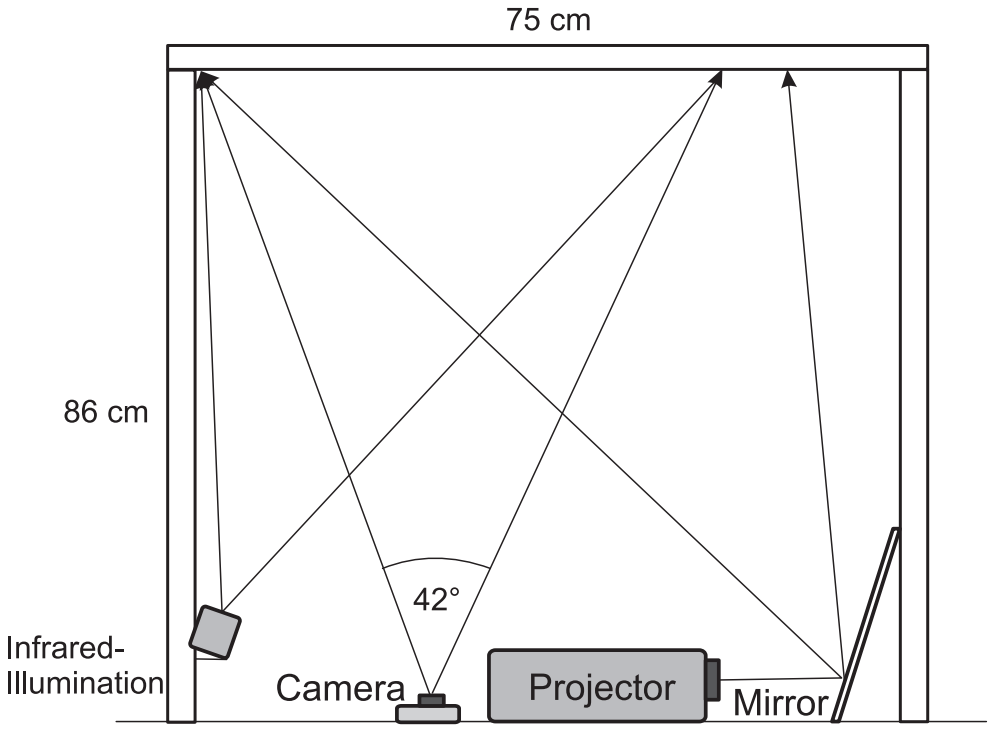


Figure 26: Table interface construction

14.1 Table construction

The setup of the table depends on both the installation environment and on the application requirements. For the installation in public spaces the table needs to be robust and accessible. In addition as musical instrument the interface needs to be mobile and easy to assemble and disassemble.

The dimensions of the table are defined in a way to achieve convenient handling of the objects and an ideal perspective on the visual feedback presented on the table surface. Thus the construction is of the dimensions: 100 cm length, 75 cm width and 86 cm height. The surface of the table is made of semitransparent, sanded glass

in order to ensure the blurring of the surface. The blurring of the surface maintains the detection of the objects only with direct contact with the table, otherwise, with transparent surfaces, the objects are detected above the table which may lead to unpredictable detection results. In addition a semitransparent surface serves as projection screen for the projection of the visual feedback.

14.2 Camera and lens

The employed Unibrain Fire-i digital camera has a IEEE-1394a (Firewire) interface and provides a resolution of VGA 640x480 with a maximum frame rate of 30 fps. The sensor is a Sony 14" color CCD with progressive scan, which allows individual readout of the image signals from all pixels. The camera supports full frame mode since an interlaced video signal destroys the structure of fiducial symbols during motion. The lens has a focal length of 4.3 mm with a 42,25 standard horizontal view angle and f 2.0 aperture. Since the camera works in the infrared spectrum the lens needs to have no infrared filter coating. An infrared pass filter is attached between lens and sensor in order to detect only infrared images on the table surface.

14.3 Projector

The projector Benq MP522 ST, with physical resolution of XGA (1024x768), an illumination of 2000 ANSI lumen in normal mode, a diagonal image size of 0,98 - 7,62 m and a possible distance from 0,5 - 5 m is used. In order to achieve a large active surface while maintaining a relatively low table height a mirror is placed in front of the projector.

14.4 Infrared illumination

In the camera-projector system the two visual components need to operate in different spectral bands in order to avoid an interference between both components. Since the projector obviously need to work in the spectrum of visible light the object tracking works in the infrared light spectrum. Thus the surface of the table need to be illuminated with strong diffuse infrared light. Suitable light sources are infrared arrays, thus two infrared light emitters from Security-Center with an illumination distance of 10-18m, a capacity of 2.5W, a wave-length of 850nm and a protection category of IP67 (water- and dust-resistant) provide diffuse illumination of the table surface. For the

14.5 Computer

In order to match the requirements of the tangible user interface the computer was individually assembled for the implementation of the instrument. Important factors in the implementation are to enable high processing performance, graphics processing, silent performance and to derive a minimum case size, which enables the fitting of the computer inside the table construction. The computer consists of a M3A79-T Deluxe motherboard, which supports the AMD Socket AM2+ and AMD 790FX chip set, a 3 GHz Phenom II X4 processor and a 4 GB RAM, 1066 MHz and dual channel mode. The passive cooled ASUS EN3700GT, silent, HTD, 256 M/A, graphics card for 16x PCI-e slots with 256 MB graphic memory is embedded. The sound card is a RME

Digi 9636/52 with digi 9652 expanding board which enable 3 x ADAT ooptical I/O, ADAT-Sync In, SPDIF I/O and word clock I/O. The ALSA Linux driver for RME digital audio interfaces enables the application of the RME audio interface with the OS Linux ALSA driver for RME Digital Audio Cards. The components are embedded in a HTPC computer case with ATX size enabling the setup of a silent PC.

15 Measurement of the room impulse response

The measurement of the room impulse response is based on the excitation of the system using an exponential sweep [Zöl08], [Far00]. Sine sweep measurements are based on a chirp signal x_s(t) of the length T_C and an inverse signal s_{inv}(t). Both signals have to satisfy the condition:

$$x_s(t) * s_{inv}(t) = \delta(t - T_C). \tag{52}$$

The chirp signal is applied to the room and the signal inside the room y(t)=x_s(t) $*$ h(n) is recorded. The impulse response is received by the convolution of the received signal y(t) with the inverse signal s_{inv}(t):

$$y(t) * s_{inv}(t) = x_s(t) * h(n) * s_{inv}(t) = h(t - T_C). \tag{53}$$

15.1 Measurement of the decay time

The estimation of the decay rate is based on the energy decay curve (EDC) which is the squared impulse response at the time t:

$$EDC(t) = \int_t^\infty h^2(\tau) d\tau. \tag{54}$$

The EDC(t) is the total amount of signal energy remaining in the impulse response at time t. The smoothing of the decay curve is derived by applying the Schroeder integration, which is the backward integration of the impulse response h(t) over the measurement interval [0,T] and converting it to a logarithmic scale [KAM$^+$02]:

$$L(\tau) = 10log_{10} \left[\frac{\int_t^T h^2(\tau) d\tau}{\int_0^T h^2(\tau) d\tau} \right] [dB]. \tag{55}$$

In order to derive the decay time a straight line, called “regression line”, is fitted in the logarithmic decay curve [KAM$^+$02]. The regression line is derived by employing the Matlab function “polyfit”. In figure 27 the regression line, from 0 dB to -60 dB in a simulated room impulse response, is illustrated in green.

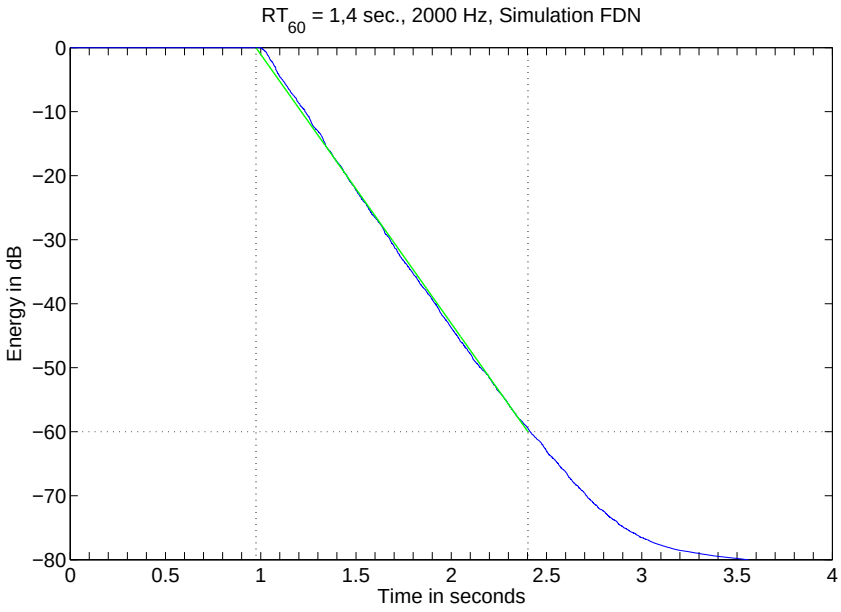


Figure 27: Decay estimation with regression line

The Schroeder frequency determines the frequency range at which the calculation of statistical parameter provide valid predictions and is calculated as follows:

$$f_{Schroeder} \approx 2000\sqrt{\frac{T_N}{V_{Room}}} \tag{56}$$

where T_N denotes the reverberation time and V_{Room} the volume of the room.

16 Simulation and analysis of the room acoustic model

In this chapter the simulation results of the implemented room acoustic model are verified. For this purpose the room acoustics of two different room sizes, in each case with three different absorption properties were simulated. In order to obtain comparative values the room acoustic prediction software $CATT - Acoustic^{TM}$ was employed to predict the room acoustics for identical room configurations.

The results of the simulation should verify the adjustment accuracy of the simulated decay time depending on the indicated room size and absorption properties. In addition predictions concerning the subjective perception of ASW and LEV, as described in chapter 6.1 should be derived. Thus, for each room configuration the reverberation time by Sabine, see chapter 4.3.3, is calculated for three different frequencies, depending on the selected absorption preset. The spatial binaural factors based on the IACF, according to Ando, as explained in chapter 7.3, and the lateral fraction (LF), in chapter 7.1 are calculated in order to provide the predictors for the above subjective spatial perceptions. In a preceding step the properties of the excita-

tion signal and the adjustment of the feedback delay network and achievable results are presented.

16.1 Generation of the impulse response (IR)

The employed excitation signal is an exponential sine sweep, generated in Matlab according to the descriptions in chapter 15 and is illustrated in figure 28. A sampling frequency fs of 48 kHz was applied with a frequency range from 10 Hz to fs/2. The generated output of the room acoustic model, introduced in chapter 13.3 was recorded in the Ambisonics encoding unit of the model for further processing and analysis in Matlab. The Ambisonics encoded signals, described in chapter 9.4, of the first and second channel (B_{00}^{+1}, B_{11}^{+1}) which are weighted with the spherical harmonic functions according to figure 6 in chapter 9.4 and the binaural signals derived by the Ambisonics based binaural sound reproduction, explained in chapter 9.6 provide the output signals for the further processing of the objective measures. The impulse response was derived, according to the descriptions in chapter 15 by the convolution of the received output signal with the inverse excitation signal.

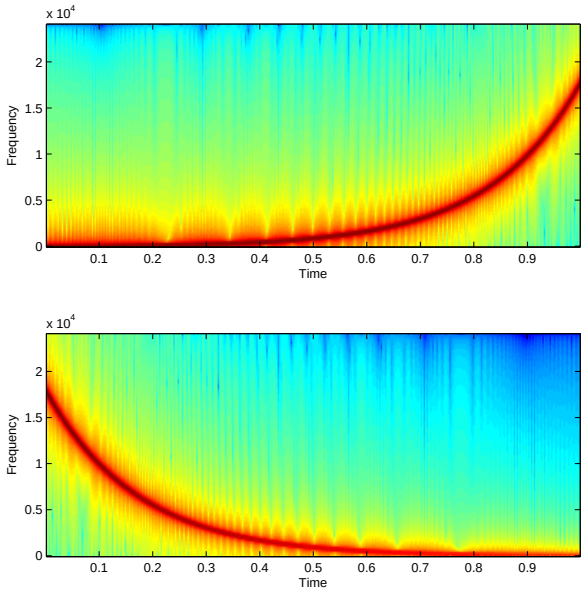


Figure 28: Exponential sine sweep and inverse sweep

16.2 Filtering of the derived IR

In order to derive frequency dependent results in the calculation of the reverberation time the impulse response is octave-band filtered, employing a first order, IIR butterworth bandpass filter. The Matlab function “filtfilt” is applied, which enables zero-phase acausal filtering of the impulse response. Thus an inducement of the phase of the signal is avoided.

16.3 Simulation of reverberation employing the FDN

According to the descriptions in chapter 13.3.2 the decay time of the simulated impulse response is adjustable for three frequency ranges. For the following simulations the low, mid and high frequency ranges were indicated by <500 Hz, 500 to 2000 Hz and >2000 Hz. In order to verify the decay behavior of the FDN, the decay time of the FDN, for the low (dt_{low}), mid (dt_{mid}) and high (dt_{high}) frequency broad band, were set to the below arbitrary values:

dt_{low} = 2,5 sec.
dt_{mid} = 1,5 sec.
dt_{high} = 1 sec.

The behavior of the FDN with the above set arbitrary values is illustrated in figure 29, which indicates the decay times for 250 Hz, 500 Hz, 1000 Hz, 2000 Hz and 4000 Hz.

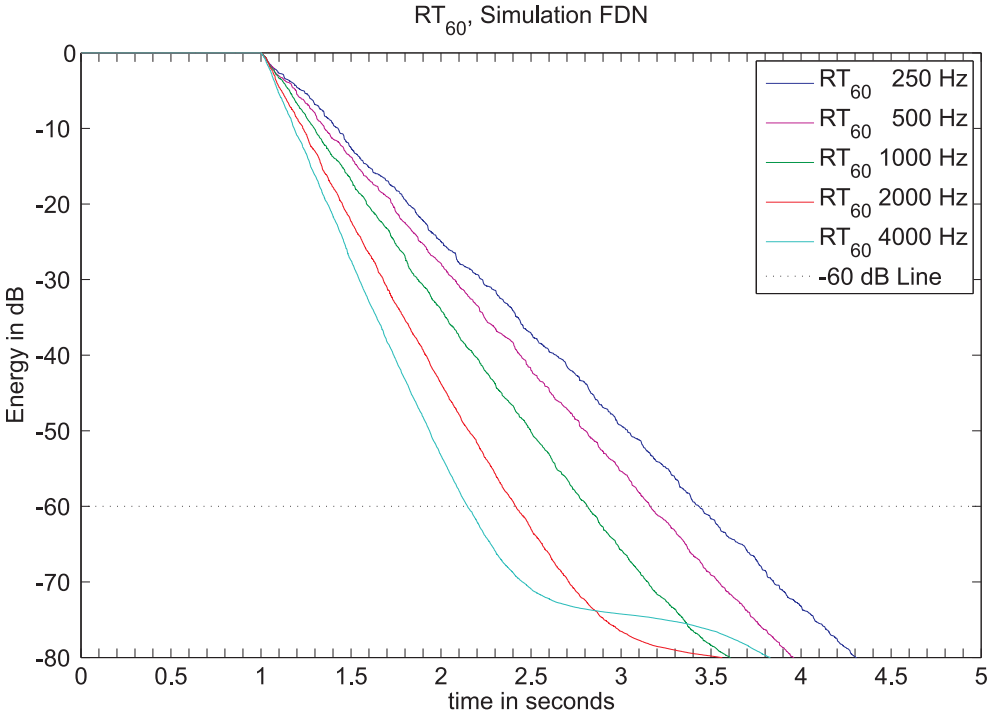


Figure 29: RT_{60} for different frequencies

16.4 Room configuration 1

For the first room configuration the room dimensions, as indicated in the following list, were set. The receiver position is located in the middle of the room. The source position is located with the distance of four meter in front of the receiver, in 0 azimuth and 0 elevation angle.

Room dimension in m	Receiver position	Source position
Breath x=10	x=5	x=5
Length y=24	y=12	y=16
Height z=8	z=1	z=1
Volume 1920 m^3		

Table 7: Room configuration 1, source and receiver position

As explained in chapter 12.3.1 each wall provides the selection of three different absorption properties. The absorption preset options and their frequency dependent degrees of sound absorption [GW05, attachment B], are listed in table 8. In order to obtain distinctive results for each preset, high mid and low absorbing materials are chosen and in the simulation for each wall, except the floor, the same preset is set. The absorption property of the floor is defined having the absorption property of an auditorium in order to compare the result with the prediction result of $CATT - Acoustic^{TM}$, where at least one auditorium plane has to be defined.

Absorbing material	Preset	250Hz	1000Hz	4000Hz
Wooden panel, 6 mm	1 wood	0,20	0,20	0,20
Wall, plastered	2 wall	0,01	0,02	0,04
Dense curtain	3 curtain	0,25	0,40	0,60
Floor		0,60	0,60	0,60

Table 8: Degrees of absorption for different materials

16.4.1 Verification of RT_{60} - room 1

Depending on the selected absorption property for each preset the the equivalent absorption area A_{equ} for three frequency bands is calculated. The room volume and the derived A_{equ} the reverberation time according to Sabine is calculated as described in chapter 4.3.3 for each frequency band. In the below table the calculated RT_{60} for each preset and frequency band is listed. The derived values further provide the settings for the decay time in three frequency band for the adjustment of the FDN.

Preset	RT_{60} 250 Hz	RT_{60} 1000 Hz	RT_{60} 4000 Hz
Wood	1	1	1
Wall	2	1,9	1.8
Curtain	0,9	0,7	0,5

Table 9: RT_{60} for different absorption presets and frequencies, V 1920 m^3

In figure 30 the derived decay curves and the reverberation time RT_{60} for 250 Hz, 1000 Hz and 4000 Hz are illustrated for absorption “preset 1, wood”. Figure 31 shows the prediction result for the global reverberation time and the mean absorption coefficient derived derived by $CATT - Acoustic^{TM}$ for equal room configuration properties. The Sabine reverberation time “SabT” is marked with the dashed line. Figure 32 and 33 compare the results for the “preset 2, wall”. In figure 34 and figure 35 the “preset 3, curtain”, are opposed. In all three cases the simulated results fit the prediction results very well except the 4kHz value of the Sabine reverberation time

for "preset 2, wall". In the following table the simulation results of the room acoustic model and the prediction results obtained by $CATT - Acoustic^{TM}$ are opposed:

Simulation	RT_{60} in s. 250 Hz	RT_{60} in s. 1000 Hz	RT_{60} in s. 4000 Hz
Wood	0,9	0,9	0,9
Wall	1,9	1,8	1,7
Curtain	0,8	0,65	0,5
$CATT - Acoustic^{TM}$			
Wood	0,95	0,95	0,82
Wall	1,74	1,84	1,23
Curtain	0,86	0,63	0,45

Table 10: RT_{60} simulation result vs. $CATT - Acoustic^{TM}$ prediction, V 1920 m^3

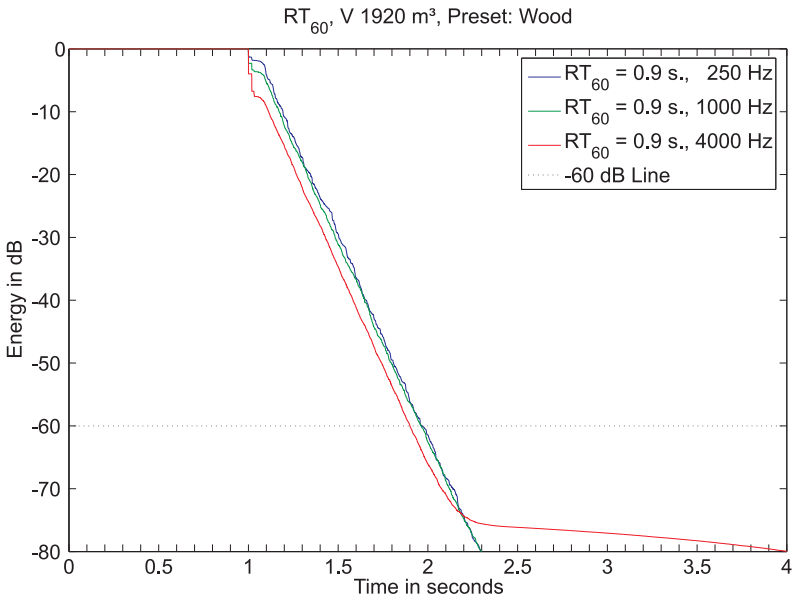
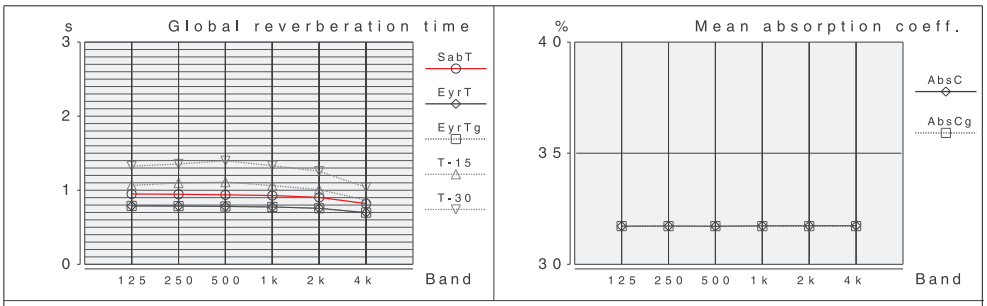


Figure 30: RT_{60}, V 1920 m^3, "preset 1, wood".



	125	250	500	1k	2k	4k	
EyrT	0,79	0,79	0,78	0,77	0,76	0,70	s
EyrTg	0,79	0,79	0,78	0,77	0,76	0,70	s
SabT	0,95	0,95	0,94	0,93	0,90	0,82	s
T-15	1,07	1,10	1,11	1,06	1,01	0,87	s
T-30	1,32	1,35	1,40	1,33	1,26	1,04	s
AbsC	31,71	31,72	31,71	31,73	31,73	31,74	%
AbsCg	31,72	31,72	31,72	31,72	31,72	31,72	%
MFP	7,49	7,50	7,50	7,50	7,49	7,50	m
Diffs	10,05	10,08	10,06	10,07	10,04	10,03	%

Trunc	4000,0	ms
Rays	10088	(used/oct)
	0	(lost/oct)
	0	(absorbed/oct)
Angle	2,02	degrees

Figure 31: Global reverberation time, mean absorption coefficients, V 1920 m^3, "preset 1, wood".

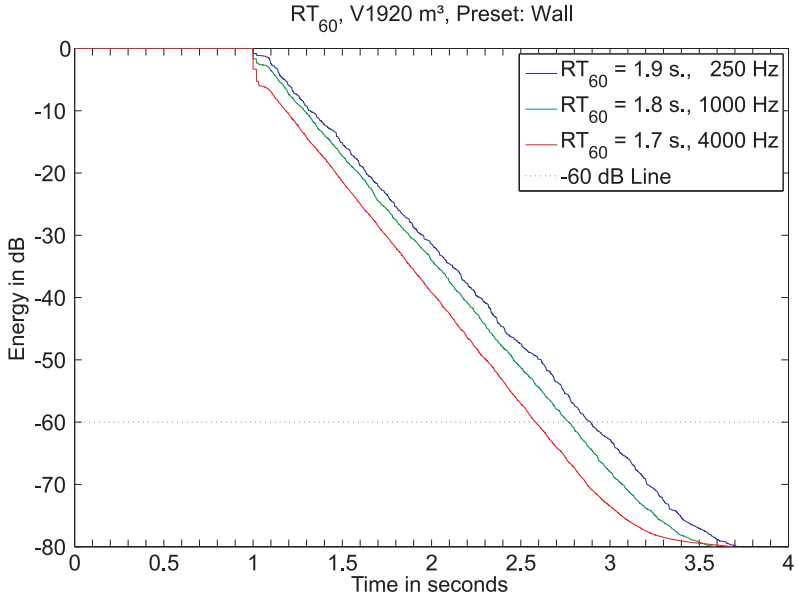
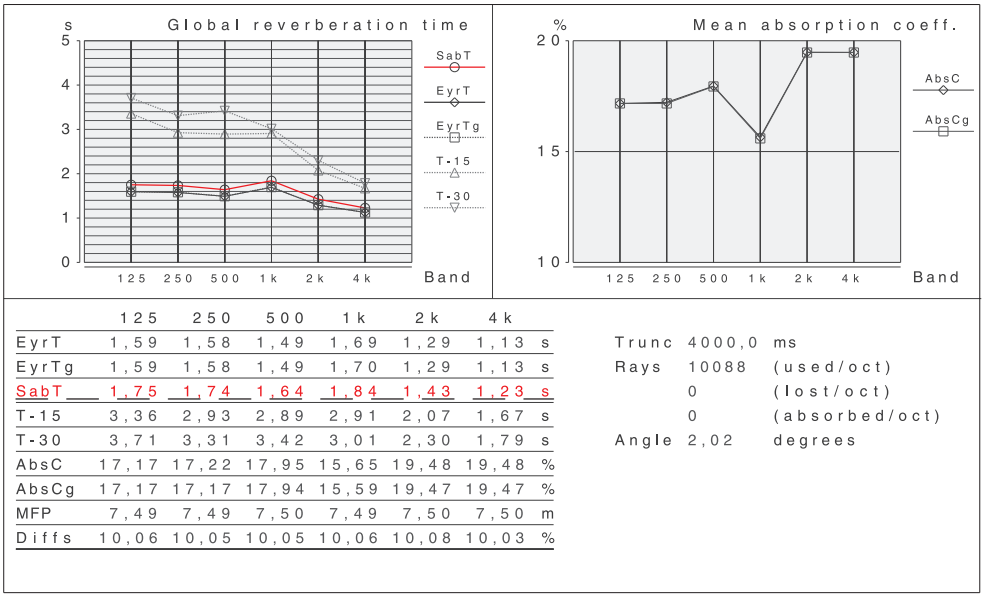


Figure 32: RT_{60}, V 1920 m^3, "preset 2, wall".



	125	250	500	1k	2k	4k	
EyrT	1,59	1,58	1,49	1,69	1,29	1,13	s
EyrTg	1,59	1,58	1,49	1,70	1,29	1,13	s
SabT	1,75	1,74	1,64	1,84	1,43	1,23	s
T-15	3,36	2,93	2,89	2,91	2,07	1,67	s
T-30	3,71	3,31	3,42	3,01	2,30	1,79	s
AbsC	17,17	17,22	17,95	15,65	19,48	19,48	%
AbsCg	17,17	17,17	17,94	15,59	19,47	19,47	%
MFP	7,49	7,49	7,50	7,49	7,50	7,50	m
Diffs	10,06	10,05	10,05	10,06	10,08	10,03	%

Trunc	4000,0	ms
Rays	10088	(used/oct)
	0	(lost/oct)
	0	(absorbed/oct)
Angle	2,02	degrees

Figure 33: Global reverberation time, mean absorption coefficients, V 1920 m^3, "preset 2, wall".

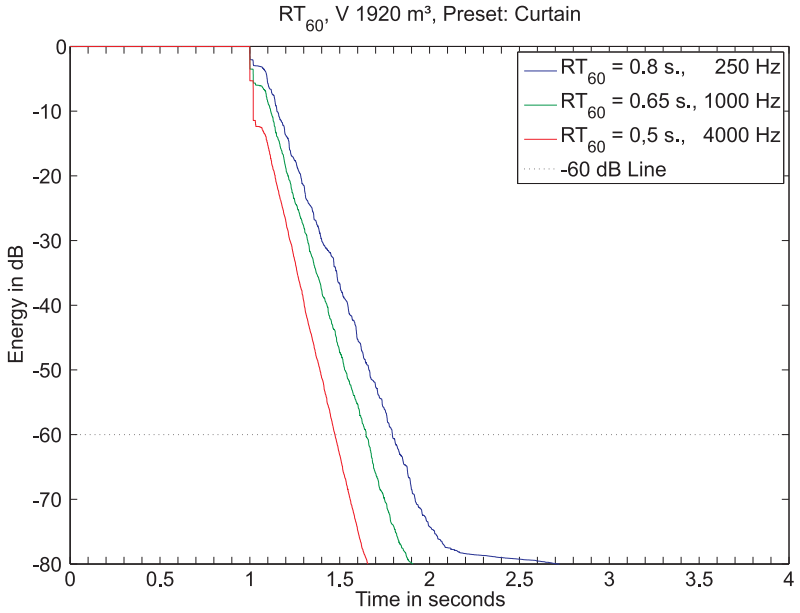
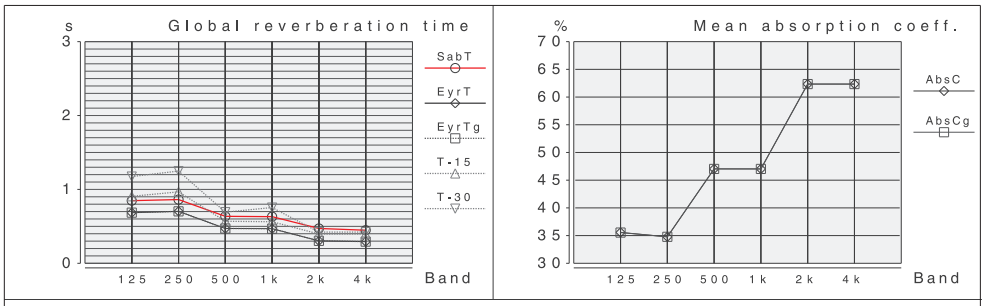


Figure 34: RT_{60}, V 1920 m^3, "preset 3, curtain".



	125	250	500	1k	2k	4k	
EyrT	0,68	0,70	0,47	0,47	0,30	0,29	s
EyrTg	0,69	0,70	0,47	0,47	0,30	0,29	s
SabT	0,85	0,86	0,64	0,63	0,47	0,45	s
T-15	0,91	0,97	0,57	0,56	0,40	0,40	s
T-30	1,17	1,25	0,69	0,75	0,42	0,42	s
AbsC	35,58	34,77	47,04	47,03	62,34	62,34	%
AbsCg	35,55	34,78	47,03	47,03	62,34	62,34	%
MFP	7,50	7,51	7,50	7,49	7,50	7,49	m
Diffs	10,01	10,04	10,07	10,06	10,05	10,05	%

Trunc	4000,0	ms
Rays	10088	(used/oct)
	0	(lost/oct)
	0	(absorbed/oct)
Angle	2,02	degrees

Figure 35: Global reverberation time, mean absorption coefficients, V 1920 m^3, "preset 3, curtain".

16.4.2 Verification binaural parameters and lateral fraction - room 1

The binaural spatial factors defined by Ando [And09] as explained in chapter 7.3 are the interaural cross correlation coefficient IACC, the interaural time difference τ_{IACC} and the $W_{IACC}(\delta)$ which is the time interval where the IACF is higher than the threshold given by ((1-δ)*IACC). According to Ando the IACC is associated with the subjective diffuseness of the sound field. Listening to a sound field with a low value for the IACC <0.15 leads to the perception of a subjectively diffuse sound. As already mentioned in chapter 7.3 the IACF is calculated, depending on the time of arrival of early and late reflections for different time intervals. Table 11 shows the calculation results of the $IACC_{Early}$, $IACC_{Late}$, $IACC_{All}$ and their corresponding interaural delay τ_{IACC}. The $IACC_{Early}$is calculated for the time interval from 0 ms to 80 ms, $IACC_{Late}$ respectively from 80 ms to 1000 ms and the $IACC_{All}$ is calculated for the time interval from 0 ms to 1000 ms. The lateral fraction (LF) values derived in the simulation is opposed by LF values derived by the $CATT - Acoustic^{TM}$ prediction.

	Preset 1, wood	Preset 2, wall	Preset 2, curtain
$IACC_{Early}$	0.96	0,95	0.97
$IACC_{Late}$	0.16	0,16	0.19
$IACC_{All}$	0,79	0,68	0.88
$\tau_{IACC(Early)}$	0 ms	0 ms	0 ms
$\tau_{IACC(Late)}$	790 μs	667 μs	810 μs
$\tau_{IACC(All)}$	0 ms	0 ms	0 ms
LF	36 %	39 %	31 %
LF $CATT - Acoustic^{TM}$	18,2 %	21,1 %	18,2 %

Table 11: IACCs and LF, V 1920 m^3

The high $IACC_{Early}$ value derived with all three preset options indicates high correlated signals at both ears in the first 80 ms. This aspect derives from the adjusted room geometries, sound source and receiver position. Since the reflection from the backward wall is arriving with the greatest delay of about 82 ms to the direct sound the diffuse reverberation occurs after 80 ms. The corresponding τ_{IACC} value equals 0 which indicates a frontal sound image corresponding to the source position of 0 azimuth and elevation angle. In contrast the low $IACC_{Late}$ values indicate a diffuse sound field after the first 80 ms and the corresponding values of τ_{IACC} are presented in the above table. The relatively high $IACC_{All}$ values predict the perception of a decreased level of LEV, since the value of IACC $< 0{,}15$ correlates with perception of a diffuse sound field. According to this results the higher values of LF indicate the perception of an increased apparent source width. Figure 36 illustrates the $IACF_{late(80-1000ms)}$ for "preset 2, wall" with the interaural delay $\tau_{IACC(Late)}$ equal 667 μs at the maximum value $IACC_{Late}$, in context to figure 4 in chapter 7.3.

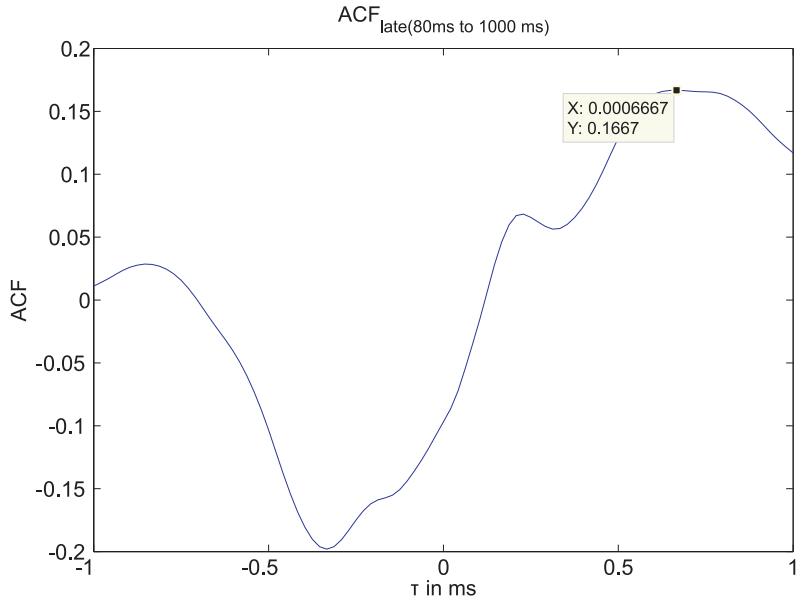


Figure 36: $IACF_{late(80-1000ms)}$, $\tau_{IACC(Late)} = 667$ μs, "preset 2, wall".

16.5 Room configuration 2

For the second room configuration the room dimensions are changed into the values, indicated in table 12. The receiver position is again located in the middle of the room. The source position is located with the distance of three meter in front of the receiver, in 0 azimuth and 0 elevation angle.

Room dimension in m	Receiver position	Source position
Breath x=8	x=4	x=4
Length y=16	y=8	y=11
Height z=5	z=1	z=1
Volume 640 m^3		

Table 12: Room configuration 2, source and receiver position

16.5.1 Verification of RT_{60} - room 2

With the same preset options and a diminished room volume of 640 m^3 the RT_{60} for the different absorption properties and frequency bands are calculated as follows:

Preset	RT_{60} sec. 250 Hz	RT_{60} sec. 1000 Hz	RT_{60} sec. 4000 Hz
Wood	0,7	0,7	0,7
Wall	1,3	1,2	1.1
Curtain	0,6	0,5	0,3

Table 13: RT_{60} for different absorption presets and frequencies, V 640 m^3

The resulting decay curves and RT_{60} for 250 Hz, 1000 Hz and 4000 Hz for the three different presets are illustrated in the figures 37, 39 and 41. The corresponding predictions of $CATT - Acoustic^{TM}$ are presented in the figures 38, 40, and 42. The simulation results of the second room configuration fit again the prediction results of $CATT - Acoustic^{TM}$. In the following table the results of the room acoustic simulation and the prediction results of $CATT - Acoustic^{TM}$ are opposed.

Simulation	RT_{60} in s. 250 Hz	RT_{60} in s. 1000 Hz	RT_{60} in s. 4000 Hz
Wood	0,65	0,65	0,64
Wall	1,1	1,1	1
Curtain	0,61	0,44	0,37
$CATT - Acoustic^{TM}$			
Wood	0,63	0,62	0,57
Wall	1,1	1,03	0,85
Curtain	0,58	0,43	0,31

Table 14: RT_{60} simulation result vs. $CATT - Acoustic^{TM}$ prediction, V 640 m^3

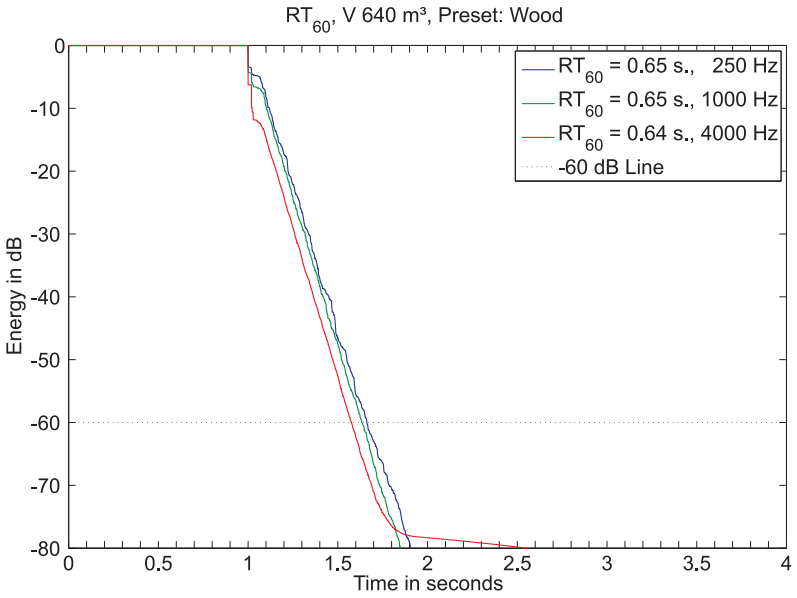
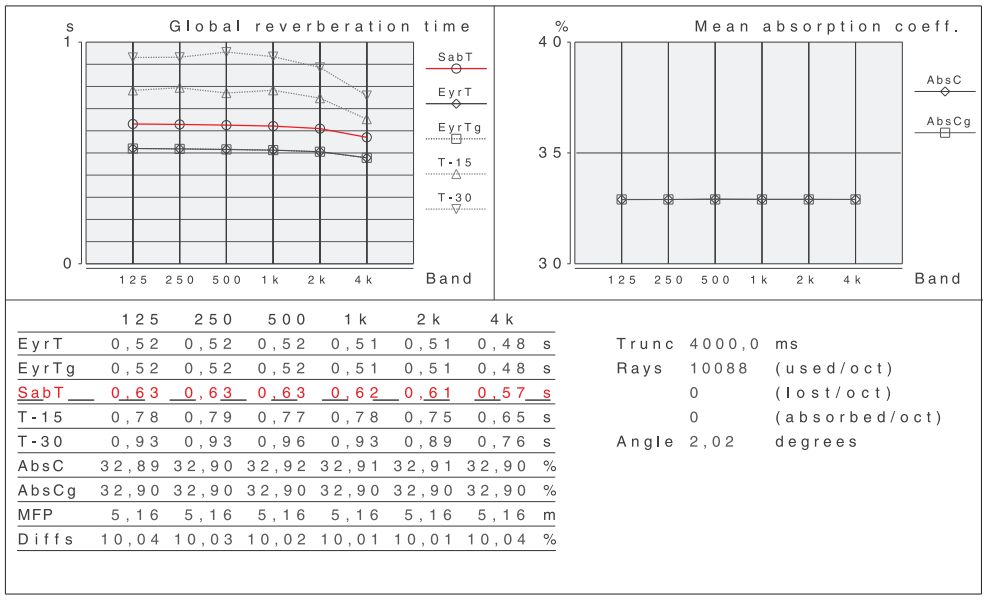


Figure 37: RT_{60} V 640 m^3, "preset 1, wood"



	125	250	500	1k	2k	4k	
EyrT	0,52	0,52	0,52	0,51	0,51	0,48	s
EyrTg	0,52	0,52	0,52	0,51	0,51	0,48	s
SabT	0,63	0,63	0,63	0,62	0,61	0,57	s
T-15	0,78	0,79	0,77	0,78	0,75	0,65	s
T-30	0,93	0,93	0,96	0,93	0,89	0,76	s
AbsC	32,89	32,90	32,92	32,91	32,91	32,90	%
AbsCg	32,90	32,90	32,90	32,90	32,90	32,90	%
MFP	5,16	5,16	5,16	5,16	5,16	5,16	m
Diffs	10,04	10,03	10,02	10,01	10,01	10,04	%

Trunc	4000,0	ms
Rays	10088	(used/oct)
	0	(lost/oct)
	0	(absorbed/oct)
Angle	2,02	degrees

Figure 38: Global reverberation time, mean absorption coefficients, V 640 m^3, "preset 1, wood".

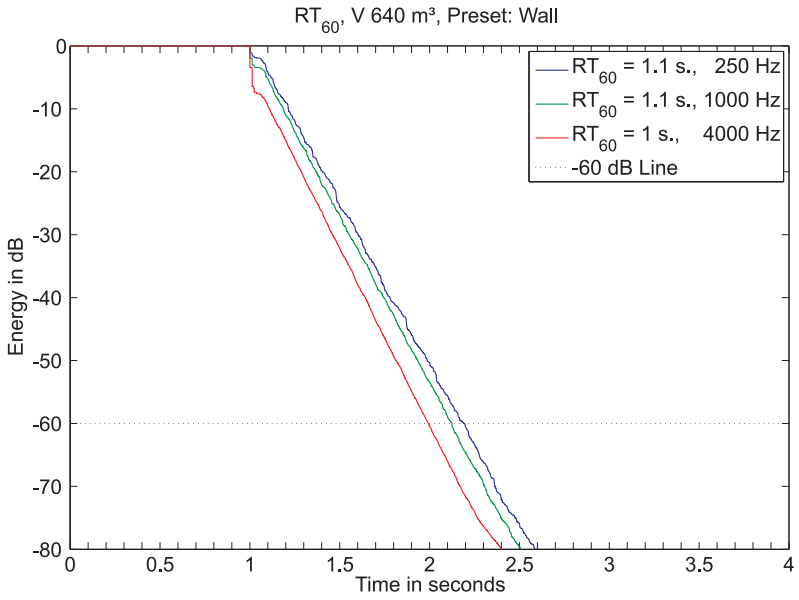
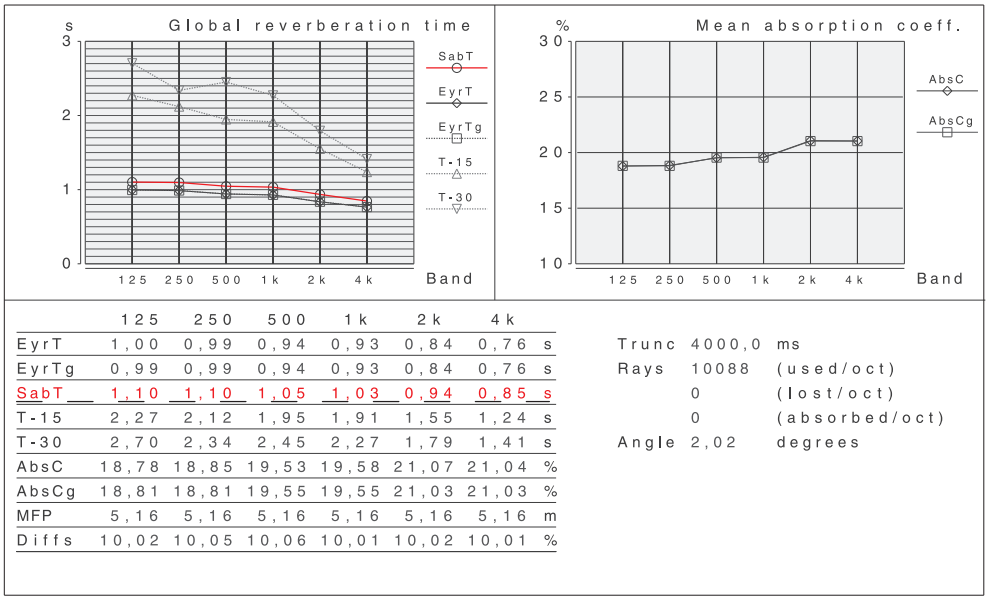


Figure 39: RT_{60} V 640 m^3, "preset 2, wall"



	125	250	500	1k	2k	4k	
EyrT	1,00	0,99	0,94	0,93	0,84	0,76	s
EyrTg	0,99	0,99	0,94	0,93	0,84	0,76	s
SabT	1,10	1,10	1,05	1,03	0,94	0,85	s
T-15	2,27	2,12	1,95	1,91	1,55	1,24	s
T-30	2,70	2,34	2,45	2,27	1,79	1,41	s
AbsC	18,78	18,85	19,53	19,58	21,07	21,04	%
AbsCg	18,81	18,81	19,55	19,55	21,03	21,03	%
MFP	5,16	5,16	5,16	5,16	5,16	5,16	m
Diffs	10,02	10,05	10,06	10,01	10,02	10,01	%

Trunc	4000,0	ms
Rays	10088	(used/oct)
	0	(lost/oct)
	0	(absorbed/oct)
Angle	2,02	degrees

Figure 40: Global reverberation time, mean absorption coefficients, V 640 m^3, "preset 2, wall".

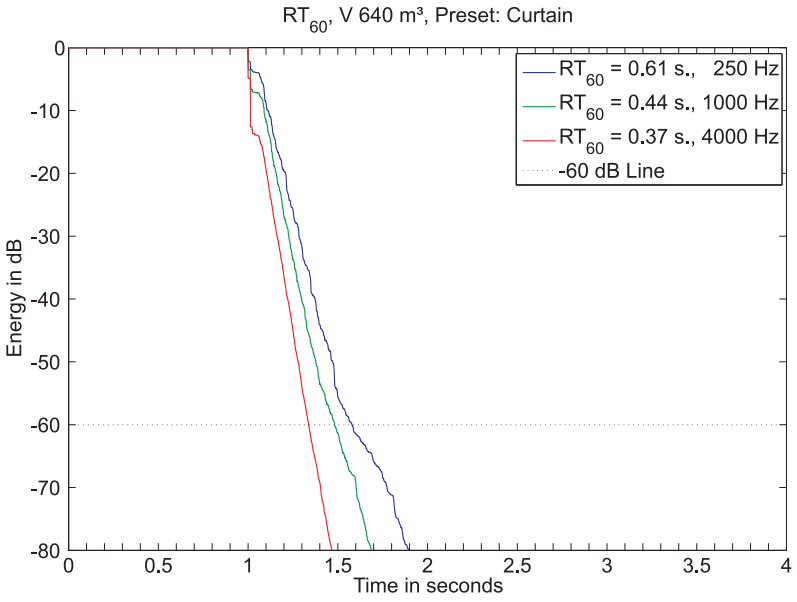
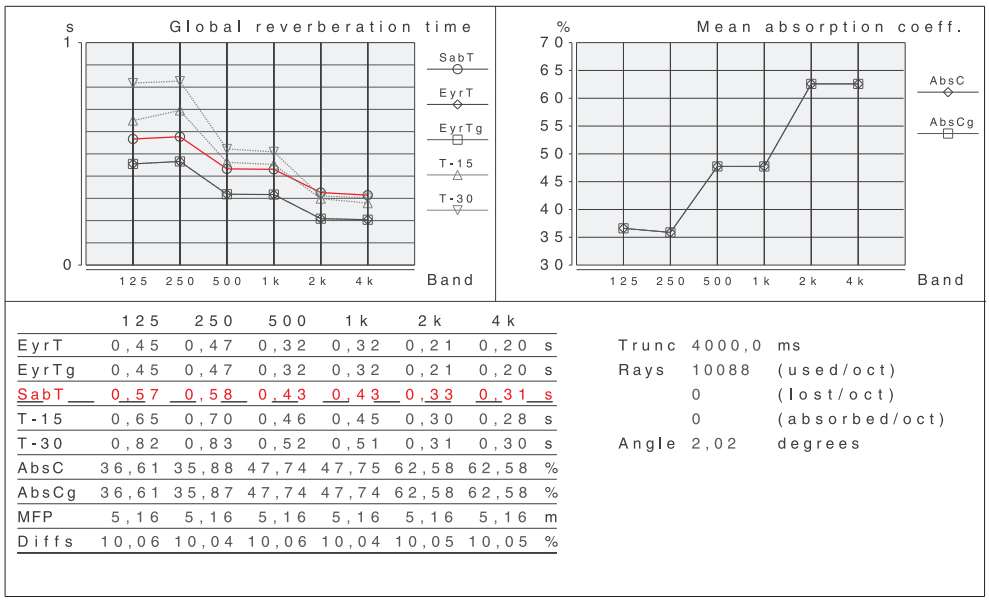


Figure 41: RT_{60} V 640 m^3, "preset 3, curtain"



	125	250	500	1k	2k	4k	
EyrT	0,45	0,47	0,32	0,32	0,21	0,20	s
EyrTg	0,45	0,47	0,32	0,32	0,21	0,20	s
SabT	0,57	0,58	0,43	0,43	0,33	0,31	s
T-15	0,65	0,70	0,46	0,45	0,30	0,28	s
T-30	0,82	0,83	0,52	0,51	0,31	0,30	s
AbsC	36,61	35,88	47,74	47,75	62,58	62,58	%
AbsCg	36,61	35,87	47,74	47,74	62,58	62,58	%
MFP	5,16	5,16	5,16	5,16	5,16	5,16	m
Diffs	10,06	10,04	10,06	10,04	10,05	10,05	%

Trunc	4000,0	ms
Rays	10088	(used/oct)
	0	(lost/oct)
	0	(absorbed/oct)
Angle	2,02	degrees

Figure 42: Global reverberation time, mean absorption coefficients, V 640 m^3, "preset 3, curtain".

16.5.2 Verification binaural parameters and lateral fraction - room 2

	Preset 1, wood	Preset 2, wall	Preset 2, curatin
$IACC_{Early}$	0,96	0,95	0,97
$IACC_{Late}$	0,15	0,17	0,21
$IACC_{All}$	0,91	0,74	0,90
$\tau_{IACC(Early)}$	0 ms	0 ms	0 ms
$\tau_{IACC(Late)}$	771 μs	708 μs	1 ms
$\tau_{IACC(All)}$	0 ms	0 ms	0 ms
LF	23 %	45 %	37 %
LF $CATT - Acoustic^{TM}$	20,2 %	25,1%	14%

Table 15: IACCs and LF, V 640 m^3

The results of the calculation of the IACCs are similar to the results of the first room configuration, whereas the higher values of $IACC_{All}$ are due to the smaller room configurations and lead to a lower level of diffuse reverberation. Figure 43 illustrates the $IACF_{late(80-1000ms)}$ for "preset 1, wood" with the interaural delay $\tau_{IACC(Late)}$ equal 771 μs at the maximum value $IACC_{Late}$.

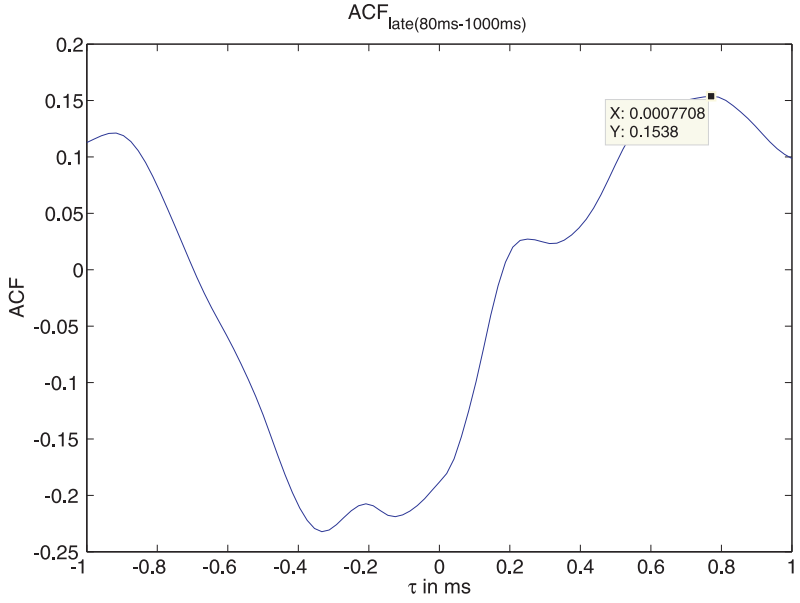


Figure 43: $IACF_{late(80-1000ms)}$, $\tau_{IACC(Late)} = 771$ μs, "preset 1, wood"

17 Conclusions

In this thesis a basic concept for the setup of a digital musical instrument playing virtual architectural space in the manner of a musical instrument was proposed. A tabletop tangible user interface (TUI) was set up and a mapping strategy providing control parameter for a room acoustic model, which further represent the musical

control parameter of the instrument, were developed. In addition a visual representation of the virtual scene and a visual feedback informing about current instrument processes was developed. In a simulation of different room settings objective measures were calculated in order to estimate the subjective perception in the virtual environment. As further steps the installation of the instrument in a public space and the evaluation of the musical interaction and instrumental output quality is desirable.

References

[AB79] Jont B. Allen and David A. Berkley. Image method for efficiently simulating small-room acoustics. *Journal Acoustical Society of America*, Am. 65(4):8, 1979.

[And09] Yoichi Ando. *Auditory and Visual Sensations*. Springer, New York Dordrecht Heidelberg London, 2009.

[Bar01] Hans-Jochen Bartsch. *Taschenbuch Mathematischer Formeln*. Fachbuchverlag Leipzig im Carl Hanser Verlag München Wien, 2001.

[BCL+08] Markus Bischof, Bettina Conradi, Peter Lachenmaier, Kai Linde, Max Meier, Philipp Pötzl, and Elisabeth Andre. Xenakis: combining tangible interaction with probability-based musical composition. *Proc. 2nd Int. Conference on Tangible and Embedded Interaction*, pages 121–124, 2008.

[BDR99] Douglas Brungart, Nathaniel Durlach, and Wiliam Rabinowitz. Auditory localization of nearby sources. ii. localization of a broadband source. *Journal Acoustical Society of America*, 106:1956–68, 1999.

[BK05] Ross Bencina and Martin Kaltenbrunner. The design and evolution of fiducials for the reactivision system. *Proc. 3rd International Conf. on Generative Systems in the Electronic Arts*, 2005.

[Bla83] Jens Blauert. *Spatial Hearing, The Psychophysics of Human Sound Localisation*. The MIT Press, Cambridge, Massachusetts, 1983.

[Bla05] Jens Blauert. *Communication Acoustics*. Springer-Verlag, Berlin Heidelberg, Ruhr-University Bochum, Germany, 2005.

[Bor84] Jeffrey Borish. Extension of the image model to arbitrary polyhedra. *Journal Acoustical Society of America*, 75:1827 – 1836, 1984.

[BZ06] Søren Bech and Nick Zacharov. *Perceptual Audio EvaluationTheory, Method and Application*. John Wiley & Sons Ltd, England, 2006.

[Cho06] Virtual Choreographer. Virtual choreographer tutorials. Website, 2006. Available online at `http://virchor.sourceforge.net/siteTVC/`; visited on March 7th 2010.

[CSR03] E. Costanza, S. B. Shelley, and J. A. Robinson. D-touch: A consumer-grade tangible interface module and musical applications. *Proc. Conf. Human-Computer Interaction (HCI03), Bath, UK*, 2003.

[Dan00] Jerome Daniel. *Représentation de champs acoustiques, application a la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia*. PhD thesis, Universite Paris, 2000.

[Dav06] Paul Davis. Jack, connecting a world of audio. Website, 2001-2006. Available online at `http://jackaudio.org/`; visited on March 7th 2010.

[DB01] Sofia Dahl and Roberto Bresin. Is the player more influcenced by the auditory than the tactile feedback from the instrument. *Proc. COST G-6 Conference on Digital Audio Effects (DAFX'01), Limerick, Ireland*, 2001.

[DGG09] Dario D'Orazio, Paolo Guidorzi, and Massimo Garai. A matlab toolbox for the analysis of ando's factors. *Proc. AES 126th Int. Convention, Munich, Germany*, 50(11), 2009.

[DNM03] Jerome Daniel, Rozenn Nicol, and Sebastien Moreau. Further investigations of high order ambisonics and wavefield synthesis for holophonic sound imaging. *Proc. AES 114th Int. Convention, Amsterdam, The Netherlands*, 2003.

[Far00] Angelo Farina. Simultaneous measurement of impulse response and distortion with a swept-sine technique. In *Proc. AES 108th Int. Convention, Paris, France*, pages 18–22, 2000.

[FCE+98] Thomas Funkhouser, Ingrid Carlbom, Gary Elko, Gopal Pingali, Mohan Sondhi, and Jim West. A beam tracing approach to acoustic modeling for interactive virtual environments. In *SIGGRAPH '98: Proc. of the 25th annual conference on Computer graphics and interactive techniques*, pages 21–32, New York, NY, USA, 1998. ACM.

[FIB95] George Fithmaurice, Hiroshi Ishii, and William Buxton. Bricks: laying the foundations for graspable user interfaces. *Conf. on Human Factors in Computing Systems*, pages 442–449, 1995.

[FWSB07] Clifton Forlines, Daniel Wigdor, Chia Shen, and Ravin Balakrishnan. Direct-touch vs. mouse input for tabletop displays. *Proce. SIGCHI conference on Human factors in computing systems, San Jose, California, USA*, pages 647 – 656, 2007.

[Ger73] Michael Gerzon. Periphony: With-height sound reproduction. *Journal of the Audio Engineering Society*, 21:2–10, 1973.

[Gri97] David Griesinger. The psychoacoustics of apparent source width, spaciousness and envelopment in performance spaces. *Acta Acustica united with Acustica*, (4):721–731, 1997.

[Gri99] David Griesinger. Objective measures of spaciousness and envelopment. *Proc. AES 16th Int. Conference: Spatial Sound Reproduction*, (16-003), 1999.

[Gro] NUI Group. Touchlib:a multi-touch development kit. Website. Available online at `http://nuigroup.com/touchlib/`; visited on March 7th 2010.

[Gro96] Anoto Group. Anoto development and design tools. Website, 1996. Available online at `http://www.anoto.com/about-anoto-1.aspx`; visited on March 7th 2010.

[GW05] Gerhard Graber and Werner Weselak. Skript zur Vorlesung Raumakustik. Technical report, Institut für Breitbandkommunikation, Technische Universität Graz, 2005.

[JCry] Jean-Marc Jot and Antoine Chaigne. Digital delay networks for designing artificial reverberators. *Proc. AES 90th Int. Convention Paris, France*, 3030:16, 1991 February.

[JGAK07] Sergi Jordà, Günter Geiger, Marcos Alonso, and Martin Kaltenbrunner. The reactable: exploring the synergy between live music performance and tabletop tangible interfaces. *In TEI'07: Proc. 1st Int. Conf. tangible and embedded interaction, New York, NY, USA*, pages 139–146, 2007.

[JKGB05] Sergi Jordà, Martin Kaltenbrunner, Günter Geiger, and Ross Bencina. The reactable. *Proc. International Computer Music Conference (2005)*, 2005.

[Jor01] Sergi Jordà. New musical interfaces and new music-making paradigms. *Proc. 2001 New Interfaces for Musical Expression Int. Conference (NIME-01). Seattle, USA.*, page 5, 2001.

[Jor02] Sergi Jordà. Improvising with computers: A personal survey (1989-2001). *J. New Music Research*, 31:1–10, 2002.

[Jor04a] Sergi Jordà. Digital instruments and players: Part i-efficiency and apprenticeship. *Proc. 2004 New Interfaces for Mucical Expression Int. Conference (NIME-04)*, page 5, 2004.

[Jor04b] Sergi Jordà. Digital instruments and players: Part ii-diversity, freedom and control. *Int. Computer Music Conference*, page 5, 2004.

[Jot97] Jean-Marc Jot. Efficient models for reverberation and distance rendering in computer music and virtual audio reality. *Proceedings 1997 International Computer Music Conference*, page 8, 1997.

[KAD09] Kenrick Kin, Maneesh Agrawala, and Tony DeRose. Determining the benefits of direct-touch, bimanual, and multifinger input on a multitouch workstation. *Proc. Graphics Interface 2009, Kelowna, British Columbia, Canada*, pages 119–124, 2009.

[Kal10] Martin Kaltenbrunner. Tangible musical interfaces. Website, 2003 - 2010. Available online at `http://modin.yuri.at/tangibles/?list=1`; visited on March 7th 2010.

[KAM+02] Matti Karjalainen, Poju Antsalo, Aki Mäkivirta, Timo Peltonen, and Vesa Välimäki. Estimation of modal decay parameters from noisy response measurements. *Journal of the Audio Engineering Society*, 50(11), 2002.

[KB07] Martin Kaltenbrunner and Ross Bencina. reactivision: A computer-vsision framework for table-based tangible interaction. *Proc. 1st international conference on Tangible and embedded interaction, Baton Rouge, Louisiana*, pages 69 – 74, 2007.

[KBBC05] Martin Kaltenbrunner, Till Bovermann, Ross Bencina, and Enrico Costanza. TUIO - a protocol for table-top tangible user interfaces. *Proc. 6th Int. Workshop on Gesture in Human-Computer Interaction and Simulation (GW 2005), Vannes, France*, 2005.

[KDS93] Mendel Kleiner, Bengt-Inge Dalenbäck, and Peter Svensson. Auralization - an overview. *Journal of the Audio Engineering Society*, 41(11):861–875, 1993.

[KJGA06] Martin Kaltenbrunner, Sergi Jordà, Günter Geiger, and Marcos Alonso. The reactable: A collaborative musical instrument. In *WETICE '06: Proceedings of the 15th IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises*, pages 406–411, Washington, DC, USA, 2006. IEEE Computer Society.

[Kut09] Heinrich Kuttruff. *Room Acoustics*. Spon Press, London and New York, 2009.

[Lab] Mitsubishi Electric Research Laboratories. Diamond touch software development kit (sdk). Website. Available online at `http://www.merl.com/projects/dtsdk/`; visited on March 7th 2010.

[Luc05] Alvin Lucier. *Reflections, Interviews, Scores, Writings*. Lovely Music Ltd., 260 West Broadway New York, New York, 2005.

[MNH05] Thomas Musil, Markus Noisternig, and Robert Höldrich. A library for realtime 3d binaural sound reproduction in pure data (pd). *Proc. 8th Int. Conference on Digital Audio Effects*, 2005.

[MPH04a] Teemu Mäki-Patola and Perttu Hämäläinen. Effect of latency on playing accuracy of two gesture controlled continuous sound insruments without tactile feedback. *Proc. 7th Int. Conference on Digital Audio Effects (DAFX'04), Naples, Italy*, page 6, 2004.

[MPH04b] Teemu Mäki-Patola and Perttu Hämäläinen. Latency tolerance for gesture controlled continuous sound instrument without tactile feedback. *Scholarly Publishing Office, University of Michigan Library*, page 8, 2004.

[MRZry] Thomas Musil, Winfried Ritsch, and Johannes M Zmölnig. The cubemixer a performance-, mixing- and masteringtool. *Proc. of the LAC 2008, Cologne, Germany*, page 5, 2008 February.

[NKSS08] Markus Noisternig, Brian Katz, Samuel Siltanen, and Lauri Savioja. Framework for real-time auralization in architectural acoustics. *Acta Acustica united with Acustica*, 94:1000–1015, 2008.

[NSMH03] Markus Noisternig, Alois Sontacchi, Thomas Musil, and Robert Höldrich. A 3d ambisonic based binaural sound reproduction system. *Proc. AES 24th Int. Convention, Banff, Canada*, 2003.

[Pel01] Renato Pellegrini. Quality assessment of auditory virtual environments. *Proc. of the 2001 International Conference on Auditory Display, Espoo, Finland*, page 8, 2001.

[PRI02] James Patten, Ben Recht, and Hiroshi Ishii. Audiopad: a tag-based interface for musical performance. In *NIME '02: Proceedings of the 2002 conference on New interfaces for musical expression*, pages 1–6, Singapore, Singapore, 2002. National University of Singapore.

[Puc96] Miller Puckette. Pure data: another integrated computer music environment. *Proc. of the Second Intercollege Computer Music Concerts, Tachikawa, Japan*, page 3741, 1996.

[Pul97] Ville Pulkki. Virtual sound source positioning using vector base amplitude panning. *Journal of the Audio Engineering Society*, 45:456–466, 1997.

[Sch61] Manfred R. Schroeder. Natural sounding artificial reverberation. *Proc. AES 13th Int. Convention*, (203):5, 1961.

[SLC03] Gilbert Soulodre, Michel Lavoie, and Scott Corcross. Objective measures of listener envelopment in multichannel surround systems. *Journal of the Audio Engineering Society*, 21:826–840, 2003.

[SRA08] Sascha Spors, Rudolf Rabenstein, and Jens Ahrens. The theory of wave field synthesis revisited. *Proc. AES 124th Int. International Convention, Amsterdam, The Netherlands*, (7358), 2008.

[Tre96] Marc Treib. *Space calculated in seconds, The Philips Pavilion, Le Corbusier, Edgard Varese*. Princeton University Press, 41 William Street, Princeton, New Jersey, USA, 1996.

[Ull95] Brygg Ullmer. *Tangible Interfaces for Manipulating Aggregates of Digital Infromation*. PhD thesis, University of Illinois, 1995.

[Vor08] Michael Vorländer. *Auralization Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality*. Springer-Verlag, Berlin Heidelberg, RWTH Aachen, Institut für Technische Akustik, Aachen, Germany, 2008.

[WD04] Marcelo Wanderley and Philippe Depalle. Gestural control of sound synthesis. *Proc. of the IEEE*, 92:632–644, 2004.

[Wei08] Stefan Weinzierl. *Handbuch der Audiotechnik*. Springer-Verlag, Berlin Heidelberg, 2008.

[WFM03] Matthew Wright, Adrean Freed, and Ali Momeni. Opensound control: state of the art 2003. *Proc. 2003 conference on New Interfaces For Musical Expression, Montreal, Quebec, Canada*, pages 153 – 160, 2003.

[Wik09] the free encyclopedia Wikipedia. Venetian polychoral style. Website, 2009. Available online at `http://en.wikipedia.org/wiki/Venetian_polychoral_style`; visited on March 7th 2010.

[Zah02] Pavel Zahorik. Assessing auditory distance perception using virtual acoustics. *Journal Acoustical Society of America*, 111:1832–46, 2002.

[ZLM+05] Vit Zouhar, Rainer Lorenz, Thomas Musil, Johannes Zmölnig, and Robert Höldrich. Hearing varèse's poème électronique inside a virtual philips pavilion. *Proceedings of ICAD 05-Eleventh Meeting of the International Conference on Auditory Display, Limerick, Ireland*, page 6, 2005.

[Zöl02] Udo Zölzer. *DAFX - Digital Audio Effects*. John Wiley & Sons Ltd, England, 2002.

[Zöl08] Udo Zölzer. *Digital Audio Signal Processing*. John Wiley & Sons Ltd, England, 2008.