Technische Universität Berlin

Doctoral Thesis

Binaural processing for the evaluation of acoustical environments

Fabian Brinkmann

audiokommunikation

Binaural processing for the evaluation of acoustical environments

https://dx.doi.org/10.14279/depositonce-8510

vorgelegt von
Fabian Brinkmann, M.A.
ORCID: 0000-0003-1905-1361

von der
Fakultät I – Geistes- und Bildungswissenschaften
Institut für Sprache und Kommunikation
Fachgebiet Audiokommunikation
der Technischen Universität Berlin

zur Erlangung des akademischen Grades
Doctor rerum naturalium (Dr. rer. nat.)

genehmigte Dissertation

Promotionsausschuss
Vorsitzender: Prof. Dr. Thorsten Roelcke
Gutachter: Prof. Dr. Stefan Weinzierl
Gutachter: Prof. Dr. Michael Vorländer
Gutachter: Prof. Dr. Christoph Pörschmann

Tag der wissenschaftlichen Aussprache
21. Mai 2019

Berlin 2019

This work is published under the CC-BY 4.0 license,
except for Chapter 3, which remains under the copyright of the IEEE,
and Chapter 4 and Appendix A, which remain under the copyright of the AES.

“Guck mal Oma, ich bin Doktor.”

Abstract

The main concern of this work is binaural processing for acoustical environments, and the evaluation of room acoustical simulations against the corresponding real sound fields. For this purpose the real sound fields and the simulations were virtualized by means of binaural synthesis – a method that aims at reproducing the sound pressure signals at the listener's ear drums. Before the comparison could be achieved, the performance of binaural synthesis was analyzed, and the input data that is required for acoustical simulations and perceptual evaluations were acquired. Results from this thesis show that most simulations can render plausible replica of the acoustic reality, while remaining perceptual differences are attributable to simplifications of the modeling algorithms and to uncertainties in collecting the input data.

Zusammenfassung

Das Hauptanliegen dieser Arbeit ist die binaurale Signalverarbeitung für akustische Umgebungen und die Evaluation raumakustischer Simulationen im Vergleich zu den entsprechenden realen Schallfeldern. Zu diesem Zweck wurden sowohl die simulierten, als auch die realen Schallfelder mittels Binauralsynthese virtualisiert – einem Verfahren, dass darauf abzielt die Schalldrucksignale am Trommelfell einer Hörerin zu reproduzieren. Vor der eigentlichen Evaluation wurde dazu die Leistungsfähigkeit der Binauralsynthese untersucht und die für die Simulation und Evaluation benötigten Eingangsdaten zusammengetragen. Die Ergebnisse dieser Arbeit zeigen, dass die meisten Simulationsverfahren eine plausible Abbildung der akustischen Realität liefern und dass verbleibende perzeptive Unterschiede zum einen vereinfachenden Annahmen der Simulationsalgorithmen zuzuweisen sind und zum anderen aus Messunsicherheiten in den Eingangsdaten resultieren.

Contents

1
Introduction

Acoustical simulation enables the computer based calculation of sound fields inside a wide range of acoustic environments such as open plan offices, lecture and concert halls, churches, stadiums, train stations, as well as entire streets or city blocks. The variety of environments that can be simulated opens a large field for the application of acoustic simulation in turn: In architectural design, it is used to improve the acoustics of existing and future buildings. In city planing, it finds use in estimating and reducing the noise exposure in residential areas. In the context of virtual and augmented reality, it offers possibilities to create spatial audio for computer games, guidance systems for visually impaired people, or educational material in museums and memorial places. Moreover, acoustical simulation is also used as a research tool, for instance to asses the effect of room acoustics on the intelligibility of speech or the performance and perception of music.

While most of these cases do not require an authentic simulation that perfectly matches the acoustics of the actual environment, a key demands towards the quality of acoustical simulations is plausibility, which demands that the simulated sound field should be in agreement with the expectation towards the acoustics of the actual environment. The plausibility, as an integral quality measure, can be interpreted as a general requirement for the validity of the simulation, and might be forgiving if some acoustic aspects are modeled less realistic than others. However, there are key aspects to each application that deserve special attention, as for example the perceived location of a sound source in a guidance system. If these aspects are well modeled, it is reasonable to assume that the simulation is able to correctly reflect changes in the acoustic environment to the extent that is necessary for the application, might it be a positional change of a source or an acoustic treatment – and clearly, all applications mentioned above rely on tracking such changes.

Previous studies that assessed the quality of room acoustical simulation algorithms mostly focused on the evaluation of room acoustical parameters of complex *real life* environments[1]. They observed that differences between parameters from different algorithms, and differences between simulated and measured parameters exceed the just noticeable difference. However, despite the fact that the per-

[1] e.g., M. Vorländer (**1995**). "International round robin on room acoustical computer simulations" in *15th International Congress on Acoustics*.

ceptual meaning of at least some of these parameters is well established – for example the early decay time and reverberation time are good predictors for the perceived reverberation – three problematic aspects are not covered by this evaluation approach. First, not all perceptual aspects are reflected by rather simple room acoustic parameters. Second, it remained unclear to what extend interactions between parameters influence perceptual quality aspects[2], and third, this approach can not explain where differences between simulations and measurements originate. An investigateion in a more controlled environment – the Bell Labs box[3] – tackled the third problematic aspect, but evaluated only one simulation algorithm. Only a single study provided publicly available reference data to enable a cross-algorithm comparison of simulation results[4]. However, the data appears to be rarely used, potentially because it comprises only a small selection of acoustic scenes.

[2] cf. S. Weinzierl and M. Vorländer (**2015**). "Room acoustical parameters as predictors of room acoustical impression: What do we know and what do we want to know?" *Acoustics Australia*.

[3] N. Tsingos, et al. (**2002**). "Validating acoustical simulations in the Bell Labs Box" *IEEE Computer Graphics and Applications*.

[4] D. Schröder, et al. (**2010**). "Open acoustic measurements for validating edge diffraction simulation methods" in *Baltic-Nordic Acoustic Meeting (BNAM)*.

The main concern of this thesis is thus the systematic evaluation of various room acoustical simulations including a focus on perceptual quality aspects. This evaluation was based on spectro-temporal comparisons and on *auralizations* of the acoustics environments, i.e. simulation results that were made audible[5]. Besides assessing the state of the art, this work also aimed at enabling users and developers of room acoustical simulation software to evaluate and improve their simulation results and algorithms in a simulation accompanying workflow. The foundation that is needed to achieve this goal is to provide (a) a public database with acoustic scenes that can be used for the evaluation, (b) a reference to which the simulation can be compared, and (c) experimental methods for the perceptual evaluation.

[5] M. Kleiner, et al. (**1993**). "Auralization – An overview" *J. Audio Eng. Soc.*

This chapter continues with introducing the basic concepts and mechanisms of spatial hearing, room acoustical simulation, and headphone based auralization in Sections 1.1 – 1.3. For the sake of brevity, this is limited to the amount that is necessary to comprehend the remainder of this thesis, and the interested reader is kindly referred to more extensive literature that is linked in the corresponding sections. The chapter closes with an outline of the thesis in Section 1.4 that details the connection between Chapters 2 – 6 and their contribution to the main objectives of this thesis.

1.1 Spatial hearing

The processing of the acoustic environment was termed *auditory scene analysis*[6]. An optical analogy of this process is illustrated in Figure 1.1, where a person tries to analyze the scene – for example the number, position, size, and speed of the objects in the water – solely by observing the water movements at two spatially separated points by the shore of a lake. In the acoustic world, the objects are sound sources that emit acoustic waves, and the spatially separated tissues that visualize the water movement are the two ear drums that are

[6] A. S. Bregman (**1994**). *Auditory scene analysis. The perceptual organization of sound*.

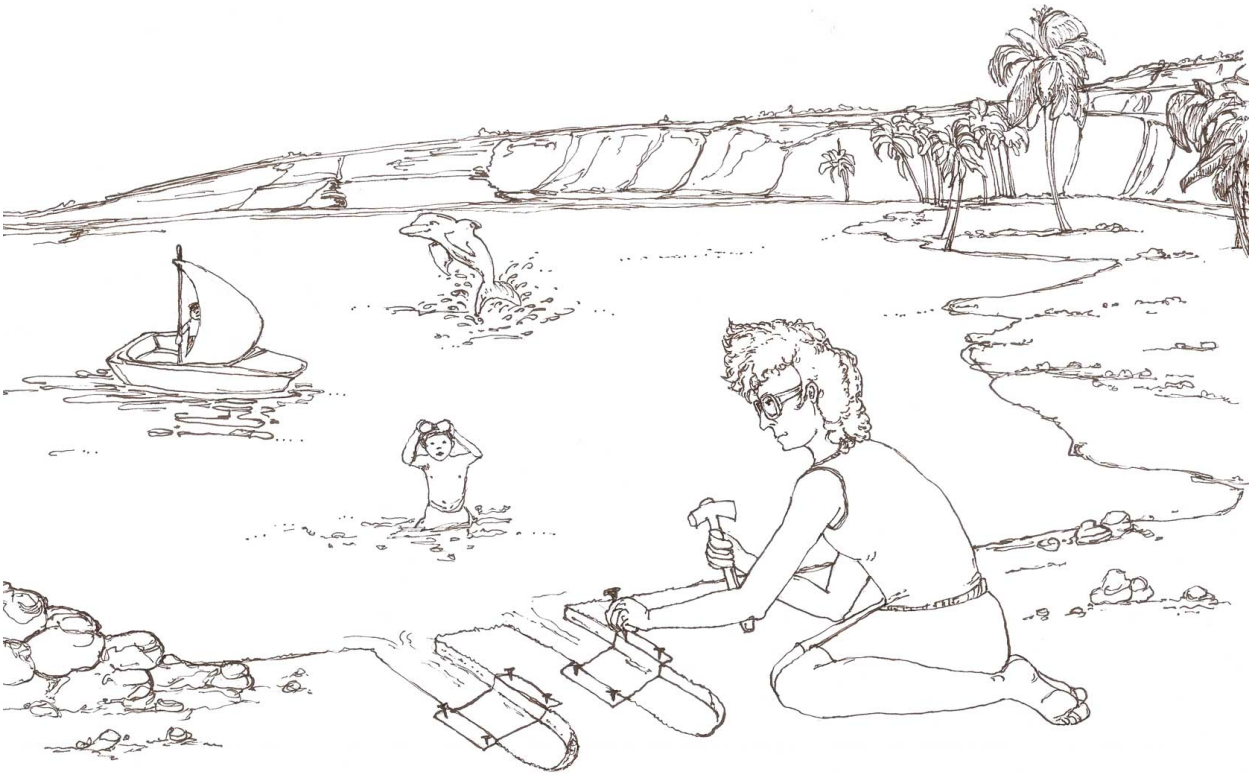


Figure 1.1: Illustration of auditory scene analysis by an optic analogy. After: A. S. Bregman (**1994**). *Auditory scene analysis. The perceptual organization of sound*, pp. 5.

separated by the head. The ear drums help to transform acoustic energy to electric impulses, which are interpreted by the auditory system to form a mental representation of the acoustic environment.

The part of auditory scene analysis that is concerned with the location of sound sources and the perception of sound in rooms is known as *spatial hearing*[7]. It exploits binaural[8] cues that stem from the spatial separation of the two ears and are derived from a comparison of the left and right ear signals, and monaural cues that originate from the listener morphology and are derived separately for each ear. Binaural cues are divided into interaural[9] time differences (ITD) that mainly stem from the spatial separation of the ears, and interaural level differences (ILD) that mainly stem from the acoustic shadow of the head and torso. The perceptually dominant time shifts cause interaural phase differences (IPD) that are evaluated below approximately 1.5 kHz, whereas level differences are larger for higher frequencies where the wave length is small compared to the head. With respect to localization, these cues are used to determine the horizontal (left/right) position of an auditory event. In contrast, the determination of the vertical (up/down) position is based on monaural spectral cues that originate from the pinnae[10]. They constitute a system of acoustic resonators and reflectors that alters the spectral content of incoming audio signals in dependence on the source position.

Additional features emerge when listening in rooms due to re-

[7] J. Blauert (**1997**). *Spatial Hearing. The psychophysics of human sound localization.*

[8] Listening with both ears (from latin *bini* – two at a time, and *auris* – ear)

[9] i.e. between ear

[10] latin *outer ears*

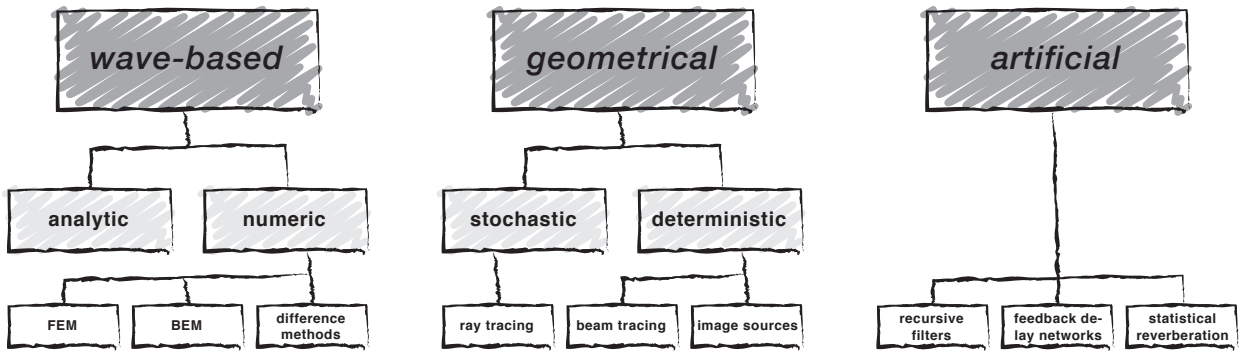


Figure 1.2: Overview of methods for room acoustical simulation. Adopted from M. Vorländer (**2008**). *Auralization. Fundamentals of acoustics, modelling, simulation, algorithms and acoustic virtual reality.*

flections from the walls and objects inside the room. Early reflections – arriving within approximately 1 ms after the direct sound – influence the perceived source position, width, and timbre. They moreover add energy that improves the intelligibility of speech and music if they arrive within about 50 ms to 80 ms after the direct sound. Late reflections, on the contrary, contribute to the sensation of envelopment by sound and do not cary information on the source position. Regardless of the arrival time, the reflections affect the temporal structure of the ear signals. The auditory system assesses this feature by means of the interaural cross correlation (IACC), which is a measure for the similarity of the ear signals in consideration of the interaural time difference[11]. Noteworthy, most auditory features are evaluated in auditory bands, which are overlapping band-pass filters of almost constant relative bandwidth, i.e., they exhibit a constant ration of the filter's center frequency and it's bandwidth.

[11] S. Klockgether and S. van de Par (**2014**). "A model for the prediction of room acoustical perception based on the just noticeable differences of spatial perception" *Acta Acust. united Ac.*

1.2 Room acoustical simulation

Room acoustical simulation aims at calculating the sound field that is evoked by an acoustic source, e.g., a loudspeaker, a singer, or a musical instrument[12,13,14]. This is done based on a simplified representation of the environment by means of a 3D room model, a description of the acoustic properties of the walls and objects inside the room, as well as a model of the source and receiver. An overview of different simulation methods that were developed throughout the past decades is given in Figure 1.2.

Wave-based simulations solve the acoustic wave equation to obtain a physically correct sound field representation. However, they depend on the correctness of the surface properties – given by the acoustic impedance – that are hard to obtain in practice. Moreover, wave based simulations are costly with respect to the required computational power and memory consumption. Although these requirements can be considerably relaxed by parallelization, wave-based methods are currently used for relatively low frequencies and simple rooms only.

Approaches based on *geometrical acoustics* are widely spread and

[12] M. Vorländer (**2008**). *Auralization. Fundamentals of acoustics, modelling, simulation, algorithms and acoustic virtual reality.*

[13] L. Savioja and U. P. Svensson (**2015**). "Overview of geometrical room acoustic modeling techniques" *J. Acoust. Soc. Am.*

[14] V. Välimäki, et al. (**2016**). "More than 50 years of artificial reverberation" in *60th In. AES Conf. DREAMS (Dereverberation and Reverberation of Audio, Music, and Speech).*

the current state of the art, at least in consumer products. They model the sound propagation based on image sources, rays, and/or beams that are radiated from the source and traced through various reflections until they reach the receiver, or their energy falls below a predefined threshold. Although complex and angle dependent acoustic impedances could be considered in theory, most algorithms use random incidence absorption and scattering coefficients to describe the surface properties, because the latter are easier to measure. While the absorption coefficient gives the amount of energy that is absorbed by a surface, the scattering coefficient is used to approximate the wave-based effect of scattering by means of the ratio of specularly reflected to diffusely scattered energy. Besides reflections on objects, the propagation of sound around them plays an important role in room acoustics. This effect, which is known as diffraction, can for example be approximated by additional image sources that are placed on edges, or a stochastic change in the direction of propagation of a ray, if it runs close to an edge. However, this is most often neglected in current simulation algorithms based on geometrical acoustics.

Artificial reverberation was originally developed in the context of sound engineering to make dry audio recordings sound more natural. They intend to reproduce key features of room acoustical environments such as the frequency dependent temporal decay rate and diffuseness. If used for room acoustical simulation – which is rather uncommon – they rely on abstract room and source representation, as for example the reverberation time, and the diffuse field transfer function.

1.3 Auralization

Auralizations can be realized based on anechoic (dry) audio content and binaural impulse responses (BIRs), which completely describe the sound propagation from a source to the listener's ears. The two can be joined by the mathematical operation of convolution that imprints the properties of the binaural impulse response onto the audio content[15]. This approach has the advantage that arbitrary combinations of audio contents and impulse responses are possible. The latter can either be acoustically measured, or simulated as described above.

[15] BIR based auralizations are also known as binaural synthesis.

Measured binaural impulse responses can be obtained from recordings of the sound pressure at the listener's blocked ear canal entrances (or at any point inside the ear canals). Two types of impulse responses are commonly distinguished: Free-field recordings from an infinite room without walls – or in practice from a room with absorbing walls – will result in so called head-related impulse responses (HRIRs). They depend on the source position, and characteristics of the source and listener, for example the size of the source, or the head width of the listener. Consequently, they contain all information required by spatial hearing, including interaural, and

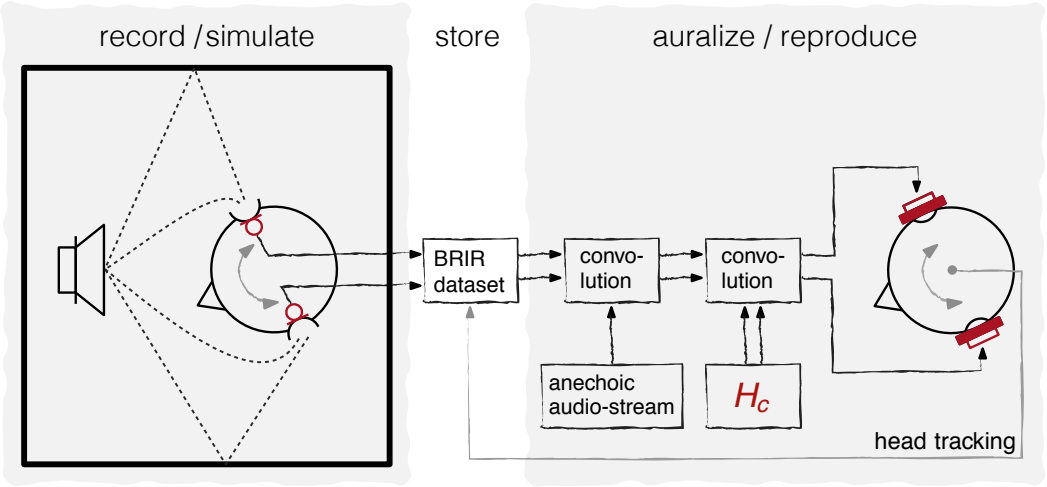


Figure 1.3: Flow diagram of headphone based dynamic binaural synthesis. Acoustic propagations paths are indicated by dashed lines, audio signals by black and control signals by gray lines. Red denotes the microphones and headphones that are used for recording and reproduction, and H_c the corresponding inverse filter.

monaural spectral cues. So called binaural room impulse responses (BRIRs) can be obtained from recordings inside rooms, and convey additional information about the room such as it's reverbarance. In the context of geometrical room acoustical simulation, BRIRs can also be interpreted as a superposition of HRIRs according to the time and direction of arrival of the direct and reflected sound. Correspondingly, geometrical modeling algorithms require a set of HRIRs as a representation of the listener to simulate BRIRs. In contrast, a 3D model of the listener's ears, head, and torso is required for wave-based simulations.

With this in mind, the fundamental idea of auralizations based on BRIRs is that a reproduction of the binaural recordings at the listener's eardrums will evoke the same auditory sensation as if the listener was present in the acoustic environment at the time and position of the recording[16]. Despite the simplicity of this idea, this makes such auraizations a valuable research tool for the perceptual evaluation of acoustic environments. With this respect, a particularly useful aspect is that it enables instant switching between auralizations of different rooms and source configurations based on stored datasets of binaural impulse responses.

For the reproduction of binaural signals loudspeakers, or headphones can be used. The latter was preferred in the current work, because it is more robust against acoustical and mechanical challenges of the reproduction environment. A signal flow graph is given in Figure 1.3 for illustration. To guarantee an unaltered reproduction, the influence of the microphones (recording device) and headphones (reproduction device) need to be equalized by a corresponding inverse filter, which is denoted by H_c in Figure 1.3. Noteworthy, listeners turn their heads in natural listening situations and when asked to evaluate spatial aspects of acoustic environments[17]. Such head movements can be accounted for in dynamic auralizations by a real time switching of the binaural impulse responses in accordance to the current head position of the listener, as illustrated

[16] H. Møller (**1992**). "Fundamentals of binaural technology" *Appl. Acoust.*

[17] C. Kim, et al. (**2013**). "Head movements made by listeners in experimental and real-life listening activities" *J. Audio Eng. Soc.*

by the arrows inside the heads in Figure 1.3. This realization of dynamic synthesis corresponds to a listener at a fixed position inside a room, that rotates is head above the torso[18]. It was shown that such head movements help to constitute externalized sound images that are perceived as being outside the head, and improve the accuracy of source localization. This is caused by dynamic (motion) cues that can be interpreted as movement induced temporal changes of the binaural cues described above.

[18] For a detailed technical description and perceptual evaluation of dynamic aspects of binaural synthesis see A. Lindau (**2014**). "Binaural resynthesis of acoustical environments. technology and perceptual evaluation" Ph.D. Thesis.

1.4 Binaural processing for the evaluation of acoustical environments

This section outlines the remainder of this thesis with a focus on the link between the chapters and their contribution to the research project – binaural processing for the evaluation of (room) acoustical environments and simulations. This cumulative thesis outlines the research project by seven selected key publications that are reprinted in Chapters 2 – 6 and Appendices A and B. The complete list of all publications that are included in this thesis is given on page 123.

To see how the previously discussed fields relate to this, the cycle of room acoustical simulation and evaluation is introduced beforehand. Figure 1.4 gives a pictographic overview starting with the input data that is required for the simulation itself. This comprises a 3D model of the scene, descriptions of the acoustic surfaces, as well as the source and receiver properties. Once the sound field is simulated, the (binaural) impulse responses can be generated generated from the software internal representation[19]. In a last step, the cycle is closed by the evaluation of the simulated impulse responses.

[19] While sound field simulation and impulse response generation are separate processes in geometrical simulations, they are combined into a single step in wave-based simulation.

The first, and most essential question for this evaluation is the question for the reference: "Against what are the simulation results compared?" Because acoustical simulations ultimately aim at recreating the acoustic reality, measured impulse responses are without a doubt the most suitable, and at the same time the most demanding reference. With this respect, gathering the input data can also be regarded as creating an image of the acoustic reality in a form that can be processed by the simulation algorithms on one hand, and serves as a reference for their evaluation on the other. From a methodical point of view, this is the question of how the reality can be operationalized. Although the measured data would not be needed to run an initial simulation, they are essential for improving the results and the underlying algorithms because the evaluation not only closes the cycle, but also initiates the next cycle after the input data were changed by the users or the algorithms were improved by the developers.

The second question is how to evaluate the simulation results. A systematic evaluation of room acoustical simulations should be able to pinpoint shortcomings of the underlying algorithms, a goal that can only be achieved by an isolated analysis of different acoustic phenomena, e.g. a single reflection on a finite plate. In such cases,

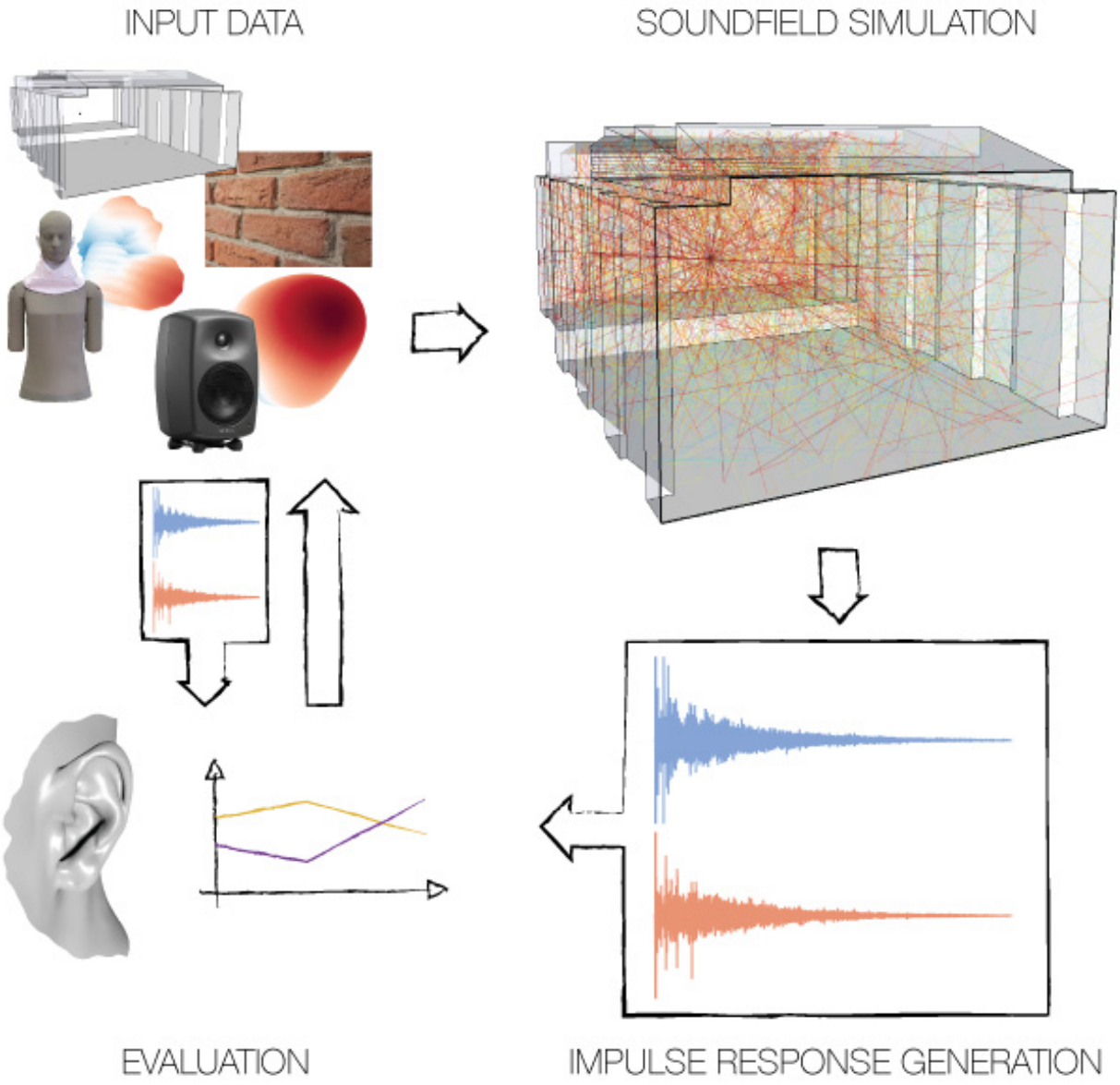


Figure 1.4: Cycle of room acoustical simulation and evaluation.

a spectro-temporal comparison of measured and simulated impulse responses – indicated by the diagram in Figure 1.4 – can be expected to already reveal valuable information for this purpose. However, room acoustical simulations are most often used for more complex environments in which case a comparison only based on physical descriptors, like room acoustical parameters, will be insufficient. Simulations of actual rooms should thus additionally be evaluated with respect to auditory perception, i.e., using listening tests (indicated by the ear in Figure 1.4). As soon as more than one acoustical environment should be perceptually evaluated, the experiments can practically not be conducted *in situ*, i.e. in the rooms of interest. This would require the availability of the rooms for a relatively long time period during the experiment, and the same listeners would have to

be available to evaluate all rooms[20]. It is thus a common choice to compare simulations and references based on auralizations of simulated and measured binaural room impulse responses. A central point within this operationalization is the listener morphology, i.e., the listener's ears, head, and torso. Necessarily, the same morphology must be used in the acoustic simulation and the reference, and ideally, this morphology corresponds to the listener that evaluates the simulation. To understand this, the *listener morphology* must be conceptually distinguished from the *listener* itself. While this may seem unintuitive at first sight, it is possible within the framework of binaural technology. Figure 1.3 shows how binaural signals can be recorded and reproduced at the listener's ears. In case the recording and reproduction are carried out with the same person, the morphology and listener are identical – which is termed *individual* binaural synthesis. We speak of *non-individual* binaural synthesis otherwise, in wich case there is a mismatch between the morphology and the listener, who literally listens through the ears of someone else in this case. In the context of binaural synthesis, the listener morphology can thus be regarded as the acoustically effective "hull", which ends at the position of the microphones that were used for the binaural recordings. The listener, in turn, can be considered the mechanico-neural processing of sound by means of the ear canal, ear drum, and the brain.

With this in mind, an individual dynamic binaural synthesis would be the ideal. This implies the use of individual head-related impulse responses for generating the binaural impulse responses from the sound field simulations, and individual binaural room impulse responses to provide a reference for the simulation results. While the first could be achieved with accelerated measurement methods[21,22], the latter is practically impossible due to the limited availability of the acoustic environments and listeners. Consequently non-individual synthesis was used throughout this study. While this will distort absolute judgements of the auditory scene, it appears reasonable to assume that relative judgements are not affected by using non-individual signals. For example, distorted spatial cues might cause errors in the perceived source location (absolute judgement). If, however, auralizations from measured and simulated data are compared against each other (relative judgement) the source locations will perceived to be identical if both auralizations are based on the same listener morphology.

Chapter 2 assesses the quality of binaural synthesis to investigate if it can be used as an alternative representation of the acoustic reality. This is of interest, because uncertainties in the measurement and dynamic reproduction of binaural impulse responses potentially degrade the quality of auralizations and make it harder to detect and evaluate differences between measurements and simulations. Conceptually, the quality of a simulation method itself can only be evaluated against the reality, and consequently, individual binaural sim-

[20] This assumes *within listener* test designs – *between listener* test designs are possible, but require more listeners.

[21] G. Enzner, et al. (**2013**). "Acquisition and representation of head-related transfer functions" in *The technology of binaural listening*, edited by J. Blauert.

[22] P. Dietrich, et al. (**2013**). "On the optimization of the multiple exponential sweep method" *J. Audio Eng. Soc.*

ulations of two loudspeakers in three rooms of varying reverberation were directly compared to the corresponding real sound fields of the same loudspeakers. In this case, individual dynamic binaural synthesis was used in order to detect even subtle differences between the simulation and reality – differences between non-individual synthesis and reality are apparent, and would have made it impossible to asses the methodical accuracy of binaural synthesis[23]. The results are promising and show that differences are inaudible in about 50% of the cases (rooms, sources, and listeners) for typical audio content like speech and music. However, the simulation could always be detected as such if listening to pink noise. Nevertheless, binaural synthesis was considered a suitable tool for evaluating room acoustical simulation, because it can be assumed that differences between simulations and measurements, as well as differences across simulation algorithms are large compared to the uncertainty related to capturing and reproducing binaural signals. This assumption is supported by a detailed analysis, which shows that differences between measured and reproduced binaural impulse responses are small in general, and of little perceptual relevance.

[23] cf. H. Møller, et al. (**1996**). "Binaural technique: Do we need individual recordings?" *J. Audio Eng. Soc.*, or A. Lindau, et al. (**2014**a). "Sensory profiling of individual and non-individual dynamic binaural synthesis using the spatial audio quality inventory" in *Forum Acusticum*.

Chapters 3 and 4 deal with the acquisition and interpolation of head-related impulse responses. To enable a dynamic binaural synthesis that allows head movements to the left and right – which are most common[24,25] – it was intended to obtain HRIRs for different head-above-torso orientations. Chapter 3 outlines the two acoustic main effects of the torso: In case the sound source, shoulder, and ear are approximately aligned, it acts as a reflector that adds energy to the direct sound, whereas it shadows the source if it is considerably below the eye level of the listener. Both effects change the temporal structure of the binaural signals, add coloration, and carry information about the height of the source, at least for certain source positions. Listening tests show that these effects are audible for speech signals and source positions which cause strong torso effects. In addition, the head orientation can always be distinguished if listening to pink noise. Following this, the interpolation of different head orientations is investigated, and it is shown that a perceptually transparent interpolation can be achieved if head orientations are available with a resolution of about 10°. HRIRs were thus measured for 11 head above torso orientations between $\pm 50^\circ$. In addition to that, HRIRs were numerically simulated to cross-validate the measured data and extrapolate the measured HRIRs at frequencies outside the working range of the loudspeakers, and source positions that could not be measured due to mechanical restriction. Results from the cross-validation show a very good agreement across measured and simulated HRIRs, and indicate a high quality of the data. Moreover, the spatial resolution of the dataset allows for a perceptually transparent and spatially continuous representations of the HRIRs.

[24] W. R. Thurlow, et al. (**1967**). "Head movements during sound localization" *J. Acoust. Soc. Am.*

[25] C. Kim, et al. (**2013**). "Head movements made by listeners in experimental and real-life listening activities" *J. Audio Eng. Soc.*

Chapter 5 is concerned with compiling the input data that is required for the room acoustical simulation on one hand, and for it's evaluation on the other. As mentioned earlier and shown in

Figure 1.4, this comprises 3D room models, information about the acoustic properties of the surfaces and materials inside the scenes, an acoustic description of the sources and receivers, and acoustically measured (binaural) impulse responses. This data was acquired in the publicly available Benchmark for Room Acoustical Simulation (BRAS[26,27]) that contains data of eleven acoustic scenes ranging from a single reflection on quasi infinite and finite plate to complex rooms. The rationale behind this was to provide simple scenes that try to isolate acoustic phenomena, as well as realistic scenarios where all acoustic principles are at work, and interact with each other. The modular database can be easily extended due it's free cultural license, which makes it appealing to the diverse community of room acoustical simulation software developers and users.

[26] L. Aspöck, et al. (**2019**). *BRAS – A Benchmark for Room Acoustical Simulation* https://dx.doi.org/10.14279/depositonce-6726.2.

[27] Please note that an earlier version of the database was used for the evaluation in Chapter 6: L. Aspöck, et al. (**2018**). *GRAS – Ground Truth for Room Acoustical Simulation* https://dx.doi.org/10.14279/depositonce-6726.

Finally, Chapter 6 details a systematic evaluation of six room acoustical simulation software packages building upon the findings and data from Chapters 2 – 5. To best investigate differences between simulation algorithms, a blind evaluation was conducted, where the input data had to be used without any changes, thereby assuming that fitting the input data – e.g., to match simulated and measured room acoustical parameters – would partly account for weaknesses of the simulation algorithms and decrease differences between them. The evaluation used spectro-temporal comparisons of measured and simulated impulse responses in case of the simple scenes, and room acoustical parameters as well as auralizations of measured and simulated binaural room impulse responses in case of the complex scenes. Moreover the perceptual evaluation comprises the two integral quality measures *authenticity* and *plausibility*, as well as a detailed analysis based on the Spatial Audio Quality Inventory (SAQI[28]). The measure for authenticity, that quantifies if any differences between the simulations and the reference are audible, was previously developed in Chapter 2. Results show that most simulations are plausible, whereas none is authentic, and that the largest differences occur for the tone color and source position. Moreover, no simulation algorithm proved to be superior with respect to all tested quality aspects and acoustic environments. Thus, a different software had to be used to obtain best results depending on the application, which indicates room for future improvements.

[28] A. Lindau, et al. (**2014**b). "A Spatial Audio Quality Inventory (SAQI)" *Acta Acust. united Ac.*

In Appendix A and B, open software and hardware that was essential for conducting the research above is introduced: AKtools are an open MATLAB toolbox for the acquisition, processing, and inspection of acoustic signals. Due to the modular design and well documented methods, the toolbox was essential for all studies mentioned above and is already used beyond the scope of this thesis and the Audio Communication Group. The PIRATE is a 3D-printable ear-plug that can be equipped with miniature microphones and enables reproducible individual binaural recordings, which was a key aspect for Chapter B[29].

[29] An earlier version of the ear-plug was used, which had to be manually crafted and was equipped with slightly larger microphones.

2

On the authenticity of individual dynamic binaural synthesis

Fabian Brinkmann, Alexander Lindau, and Stefan Weinzierl (**2017**), *J. Acoust. Soc. Am.*, **142**(4), 1784–1795. DOI: 10.1121/1.5005606.
(Accepted manuscript. CC-BY 4.0)

A SIMULATION that is perceptually indistinguishable from the corresponding real sound field could be termed *authentic*. Using binaural technology, such a simulation would theoretically be achieved by reconstructing the sound pressure at a listener's ears. However, inevitable errors in the measurement, rendering, and reproduction introduce audible degradations, as it has been demonstrated in previous studies for anechoic environments and static binaural simulations (fixed head orientation). The current study investigated the authenticity of individual dynamic binaural simulations for three different acoustic environments (anechoic, dry, wet) using a highly sensitive listening test design. The results show that about half of the participants failed to reliably detect any differences for a speech stimulus, whereas all participants were able to do so for pulsed pink noise. Higher detection rates were observed in the anechoic condition, compared to the reverberant spaces, while the source position had no significant effect. It is concluded that the authenticity mainly depends on how comprehensive the spectral cues are provided by the audio content, and the amount of reverberation, whereas the source position plays a minor role. This is confirmed by a broad qualitative evaluation, suggesting that remaining differences mainly affect the tone color rather than the spatial, temporal or dynamical qualities.

2.1 Introduction

Spatial hearing, i.e. the human ability to perceive three dimensional sound, relies on evaluating the sound pressure signals arriving at the two ear drums, and the monaural and binaural cues imprinted on them by the outer ears, the head, and the human torso. These cues depend on the position and orientation of the sound source and the listener in interaction with the properties of the sur-

rounding acoustical environment[1]. Binaural synthesis exploits these principles by reconstructing the pressure signals at a listener's ears, based on the measurement or the simulation of binaural impulse responses and a subsequent convolution with anechoic audio content[2]. If the electroacoustic signal chain (microphones, headphones) could be perfectly linearized, and if there were no measurement errors, this should result in an exact copy of the corresponding binaural sound events[3]. Early binaural simulations were mostly static, i.e. did not account for the listener's head orientation. It was shown, however, that head movements are important for sound source localization[4], improve localization accuracy[5], aid externalization[6] and are naturally used when attending concerts, playing video games, or judging perceptual qualities such as source width and envelopment[7]. This fostered the development of dynamic binaural synthesis, where binaural impulse responses are exchanged according to the listener's position and head orientation in real-time.

The time-variant nature, however, poses additional challenges on binaural signal acquisition and processing as it requires an imperceivable round-trip system latency[8], a perceptually transparent spatial discretization of the impulse response dataset [9], and suitable approaches for interpolation during head movements of the listener[10,11].

While each of these steps for signal acquisition and processing can be evaluated individually, it is not straightforward how to evaluate the entire signal chain of dynamic binaural synthesis in a comprehensive way. For this purpose, the *plausibility* and the *authenticity* of virtual acoustic environments were proposed as overall criteria for the simulated acoustical scene as well as for the quality of the systems they are generated with. While the plausibility of a simulation refers to the agreement with the listener's expectation towards a corresponding real event (agreement to an internal reference)[12], the authenticity refers to the perceptual identity with an explicitly presented real event (agreement to an external reference, Blauert (1997)[13], p. 373). Even non-individual dynamic binaural simulations recorded with a dummy head have been shown to provide plausible simulations[14,15]. The involved participants, nevertheless, always reported audible differences, even if these did not help them to identify reality or simulation as such.

At least four empirical studies were concerned with the authenticity of binaural synthesis [16,17,18,19]. In all cases, the differences between reality and simulation were audible, even if the detection rates exceeded the guessing rate only slightly (depending on the audio content, listener expertise, and the experimental setup). All of these studies were conducted as static simulations, while the authenticity of *dynamic* binaural synthesis has not been assessed before. Moreover, previous studies were restricted to anechoic environments, and the results were always cumulated across participants and test conditions, neglecting the potential differences in the individual performance of participants and effects related to audio content or to the

[1] J. Blauert (**1997**). *Spatial Hearing. The psychophysics of human sound localization*.

[2] H. Møller (**1992**). "Fundamentals of binaural technology" *Appl. Acoust.*

[3] F. L. Wightman and D. J. Kistler (**1989**a). "Headphone simulation of free field listening. I: Stimulus synthesis" *J. Acoust. Soc. Am.*

[4] W. R. Thurlow, et al. (**1967**). "Head movements during sound localization" *J. Acoust. Soc. Am.*

[5] K. I. McAnally and R. L. Martin (**2014**). "Sound localization with head movement: implications for 3-d audio displays" *Frontiers in Neuroscience*.

[6] E. Hendrickx, et al. (**2017**). "Influence of head tracking on the externalization of speech stimuli for non-individualized binaural synthesis" *J. Acoust. Soc. Am.*

[7] C. Kim, et al. (**2013**). "Head movements made by listeners in experimental and real-life listening activities" *J. Audio Eng. Soc.*

[8] D. S. Brungart, et al. (**2005**). "The detectability of headtracker latency in virtual audio displays" in *Eleventh Meeting of the International Conference on Auditory Display (ICAD)*.

[9] A. Lindau and S. Weinzierl (**2009**). "On the spatial resolution of virtual acoustic environments for head movements on horizontal, vertical and lateral direction" in *EAA Symposium on Auralization*.

[10] E. H. A. Langendijk and A. W. Bronkhorst (**2000**). "Fidelity of three-dimensional-sound reproduction using a virtual auditory display" *J. Acoust. Soc. Am.*

[11] A. Lindau, et al. (**2010**). "Individualization of dynamic binaural synthesis by real time manipulation of the ITD" in *128th AES Convention, Convention Paper*.

[12] A. Lindau and S. Weinzierl (**2012**). "Assessing the plausibility of virtual acoustic environments" *Acta Acust. united Ac.*

[13] J. Blauert (**1997**). *Spatial Hearing. The psychophysics of human sound localization*.

[14] A. Lindau and S. Weinzierl (**2012**). "Assessing the plausibility of virtual acoustic environments" *Acta Acust. united Ac.*

[15] C. Pike, et al. (**2014**). "Assessing the plausibility of non-individualised dynamic binaural synthesis in a small room" in *AES 55th International Conference*.

[16] E. H. A. Langendijk and A. W. Bronkhorst (**2000**). "Fidelity of three-dimensional-sound reproduction using a virtual auditory display" *J. Acoust. Soc. Am.*

Source	Amount typ. (max)	Reference
head repositioning	4 (10) dB	Riederer (2004) Hiekkanen *et al.* (2009)
microphone repositioning	5 (20) dB	Lindau and Brinkmann (2012)
headphone repositioning	5 (20) dB	Møller *et al.* (1995a) Paquier and Koehl (2015)
acoustic headphone load	4 (10) dB	Møller *et al.* (1995a)
headphone presense	10 (25) dB	Langendijk and Bronkhorst (2000), Moore *et al.* (2007), Brinkmann *et al.* (2014a)
headphone compensation	1 (10) dB	Lindau and Brinkmann (2012)

Table 2.1: Sources of errors and variance in the measurement and reproduction of binaural signals. Typical, and maximum errors were either directly taken from the references, or obtained by visual inspection of corresponding figures.

spatial configuration of source and receiver. Finally, an isolated test on authenticity provides little information about specific weaknesses of the binaural simulation, which might be valuable for technical improvements. In general, the literature has largely been focused on the evaluation of localization (e.g. Wightman and Kistler[20]), while other perceptual qualities, that might also be of high relevance in the context of virtual acoustic environments, were left unstudied.

In the current study, we thus combined tests of the authenticity of an individualized dynamic binaural synthesis in both anechoic and two reverberant environments with a comprehensive qualitative evaluation of 45 perceptual attributes. We aimed at designing the authenticity test to be as sensitive as possible, in order to produce practically meaningful results already at the level of individual participants, and to investigate the influence of room acoustical conditions, different source-receiver configurations, and the audio content.

The quality assessment of dynamic binaural synthesis with respect to authenticity as the strictest possible criterion is not only relevant to evaluate the performance and to identify potential shortcomings of binaural technology itself: It seems to become standard practice also to evaluate loudspeaker based reproduction systems by using a binaurally transcoded representation of the corresponding channels. This comprises the evaluation of mono/stereo loudspeaker setups[21,22] as well as loudspeaker arrays driven by sound field synthesis techniques such as wave field synthesis[23] or higher order ambisonics [24]. For this purpose, only an authentic simulation, including a natural interaction with the listener's head movements and a representation of the surrounding spatial environment, can provide a reliable and transparent reference for the perceptual evaluation of these techniques.

[17] A. H. Moore, et al. (**2010**). "An initial validation of individualised crosstalk cancellation filters for binaural perceptual experiments" *J. Audio Eng. Soc.*

[18] B. Masiero (**2012**). "Individualized binaural technology. measurement, equalization and perceptual evaluation" Doctoral Thesis.

[19] J. Oberem, et al. (**2016**). "Experiments on authenticity and plausibility of binaural reproduction via headphones employing different recording methods" *Appl. Acoust.*

[20] F. L. Wightman and D. J. Kistler (**1989**b). "Headphone simulation of free field listening. II: Psychological validation" *J. Acoust. Soc. Am.*

[21] F. L. Wightman and D. J. Kistler (**1989**b). "Headphone simulation of free field listening. II: Psychological validation" *J. Acoust. Soc. Am.*

[22] T. Hiekkanen, et al. (**2009**). "Virtualized listening tests for loudspeakers" *J. Audio Eng. Soc.*

[23] H. Wierstorf (**2014**). "Perceptual assessment of sound field synthesis" Doctoral Thesis.

[24] M. Frank (**2013**). "Phantom sources using multiple loudspeakers in the horizontal plane" Doctoral Thesis.

2.2 *Method*

Many sources of error might occur during measuring and reproducing binaural signals at the listener's ears. An inspection of errors that are relevant in the context of this study is given in Table 2.1. It shows that each of them alone can already produce potentially audible artifacts, and has to be carefully controlled if aiming at an authentic simulation. We will thus discuss these error sources and possibilities to avoid them, before we outline the setup and methods used for perceptual testing in the following.

Hiekkanen *et al.*[25] reported that head movements of 1 cm to the side and azimuthal head-above-torso rotations of 2.5° already produce audible differences in binaural transfer functions. To account for this, previous studies used some kind of head rest to restrict the participants' head position, and monitored the participants' head position with optic or magnetic tracking systems[26,27]. Throughout the study, Masiero (2012)[28] allowed for head movements between ±1 cm translation, and ±2° rotation, respectively. In the current study, we allowed tolerances of ±1 cm, and ±0.5° during the measurement of the binaural transfer functions.

Similar errors are induced by repositioning the microphones[29], or headphones[30] which was shown to be audible even for naïve listeners in the case of headphone repositioning[31]. In the context of this study such problems can be avoided, if the microphones are kept in position while measuring the binaural transfer functions of loudspeakers and headphones, and if the headphones are worn during the entire experiment.

The presence of headphones, however, influences sound field at the listener's ears because they act as an obstacle to sound arriving from the outside, as well as sound being reflected from the listener's head. This causes distortions in the magnitude and phase spectra[32,33,34], as well as changes in the acoustic load seen from inside the ear canal (free air equivalent coupling criterium[35]). To avoid this, Langendijk and Bronkhorst[36] used small extraaural earphones with a limited band width, while Moore *et al.*[37] had cross-talk cancelled transaural loudspeakers for binaural signal reproduction. We used extraaural headphones with full band width, whose influence on external sound fields are comparable to the earphones used by Langendijk and Bronkhorst[38].

Moreover, headphone transfer functions (HpTFs) show considerable distortions that need to be compensated by means of inverse filters, which are typically designed by frequency dependent regulated inversion of the HpTFs[39]. During the filter design, it is vital to find a good balance between an exact inversion that would result in filters with undesired high gains at the frequencies of notches in the HpTFs (possibly causing audible ringing artifacts), and too much regulation which is likely to cause high frequency damping[40]. To assure this, we applied regularization only at frequencies where notches in the HpTFs occurred.

[25] T. Hiekkanen, et al. (**2009**). "Virtualized listening tests for loudspeakers" *J. Audio Eng. Soc.*

[26] A. H. Moore, et al. (**2010**). "An initial validation of individualised crosstalk cancellation filters for binaural perceptual experiments" *J. Audio Eng. Soc.*

[27] B. Masiero (**2012**). "Individualized binaural technology. measurement, equalization and perceptual evaluation" Doctoral Thesis.

[28] B. Masiero (**2012**). "Individualized binaural technology. measurement, equalization and perceptual evaluation" Doctoral Thesis

[29] A. Lindau and F. Brinkmann (**2012**). "Perceptual evaluation of headphone compensation in binaural synthesis based on non-individual recordings" *J. Audio Eng. Soc.*

[30] H. Møller, et al. (**1995**a). "Transfer characteristics of headphones measured on human ears" *J. Audio Eng. Soc.*

[31] M. Paquier and V. Koehl (**2015**). "Discriminability of the placement of supraaural and circumaural headphones" *Appl. Acoust.*

[32] E. H. A. Langendijk and A. W. Bronkhorst (**2000**). "Fidelity of three-dimensional-sound reproduction using a virtual auditory display" *J. Acoust. Soc. Am.*

[33] A. H. Moore, et al. (**2007**). "Headphone transparification: A novel method for investigating the externalisation of binaural sounds" in *123rd AES Convention, Convention Paper 7166*.

[34] F. Brinkmann, et al. (**2014**). "Assessing the authenticity of individual dynamic binaural synthesis" in *Proc. of the EAA Joint Symposium on Auralization and Ambisonics*.

[35] H. Møller, et al. (**1995**a). "Transfer characteristics of headphones measured on human ears" *J. Audio Eng. Soc.*

[36] E. H. A. Langendijk and A. W. Bronkhorst (**2000**). "Fidelity of three-dimensional-sound reproduction using a virtual auditory display" *J. Acoust. Soc. Am.*

[37] A. H. Moore, et al. (**2010**). "An initial validation of individualised crosstalk cancellation filters for binaural perceptual experiments" *J. Audio Eng. Soc.*

[38] F. Brinkmann, et al. (**2014**). "Assessing the authenticity of individual dynamic binaural synthesis" in *Proc. of the EAA Joint Symposium on Auralization and Ambisonics*.

[39] S. G. Norcross, et al. (**2006**). "Inverse filtering design using a minimal phase target function from regularization" in *121th AES Convention, Convention Paper 6929*.

Note that in the case of authenticity, the room and the loudspeakers are considered to be a part of the experiment. This is in contrast to HRTF measurements, where their influence should be removed from the measured transfer functions by means of post-processing, and thus become additional sources of error[41].

[40] A. Lindau and F. Brinkmann (**2012**). "Perceptual evaluation of headphone compensation in binaural synthesis based on non-individual recordings" *J. Audio Eng. Soc.*

[41] A. Andreopoulou, et al. (**2015**). "Inter-laboratory round robin HRTF measurement comparison" *IEEE J. Sel. Topics Signal Process.*

2.2.1 *Experimental setup*

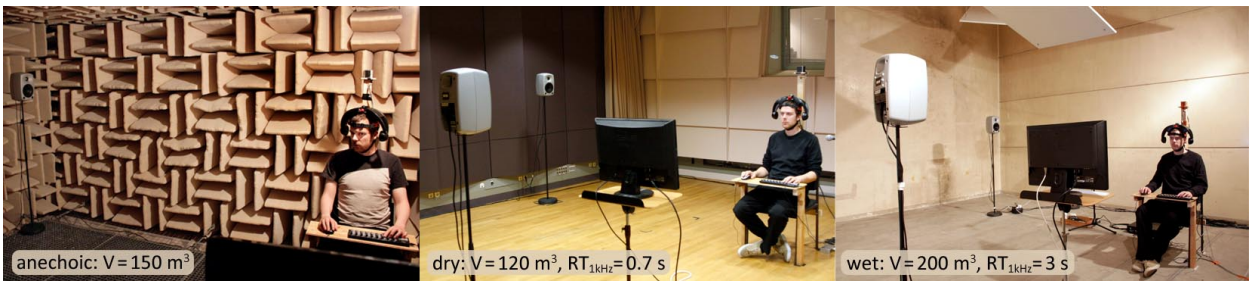


Figure 2.1: Listening test setup in the anechoic, dry, and wet test environment.

The listening tests were conducted in the anechoic chamber and the recording studio of the State Institute for Music Research, and in the reverberation chamber of TU Berlin (Figure 2.1). The three rooms are of comparable volume but exhibit large differences in reverberation time. To limit the duration of the experiment to a practical amount, the reverberation time of the wet room was reduced from 6.7 to 3 s at 1 kHz using 1.44 m^3 porous absorber. Participants were seated on a chair equipped with a height and depth adjustable neck rest, and a small table providing an arm-rest and space for placing the MIDI interface used throughout the test (Korg nanoKONTROL). An LCD screen was used as visual interface and placed 2 m in front of the participants at eye level.

Two active near-field monitors (Genelec 8030a) were placed in front and to the right of the participants at a distance of 3 m and a height of 1.56 m, corresponding to source positions of 0° and 90° azimuth, and 8° elevation. The height was adjusted so that the direct sound path was not blocked by the LCD screen. The source positions were chosen to represent the most relevant use case of a frontal source, as well as the potentially critical case of a lateral source, where the signal to noise ratio decreases due to shadowing of the head. With a loudspeaker directivity index of ca. 5 dB at 1 kHz[42] and corresponding critical distances of 1.3 m (dry) and 0.8 m (wet), the source positions result in slightly emphasized diffuse field components in the reverberant environments.

For reproducing the binaural signals, low-noise DSP-driven amplifiers and extraaural headphones were used, which were designed to exhibit minimal influence on sound fields arriving from external sources while providing full audio bandwidth (BKsystem[43]). To allow for an instantaneous switching between the binaural simulation and the corresponding real sound field, the headphones were worn during the entire listening test, i.e. also during the binaural measure-

[42] J. G. Tylka, et al. (**2015**). "A database of loudspeaker polar radiation measurements" in *139th AES Convention, e-Brief 230*.

[43] V. Erbes, et al. (**2012**). "An extraaural headphone system for optimized binaural reproduction" in *Fortschritte der Akustik – DAGA 2012*.

ments. The participants' head positions were controlled using head tracking with 6 degrees of freedom (x, y, z, azimuth, elevation, lateral flexion) with a precision of 0.001 cm and 0.003°, respectively (Polhemus Patriot). A long term test of eight hours showed no noticeable drift of the tracking system.

Individual binaural transfer functions were measured at the entrance of the blocked ear canal using Knowles FG-23329 miniature electret condenser microphones flush cast into conical silicone earmolds. The molds were available in three different sizes, providing a good fit and reliable positioning for a wide range of individuals[44]. Phase differences between left and right ear microphones did not exceed $\pm 2^\circ$ to avoid audible interaural phase distortion[45].

The experiment was monitored from a separate room with talkback connection to the test environment.

[44] A. Lindau and F. Brinkmann (**2012**). "Perceptual evaluation of headphone compensation in binaural synthesis based on non-individual recordings" *J. Audio Eng. Soc.*

[45] A. W. Mills (**1958**). "On the minimum audible angle" *J. Acoust. Soc. Am.*

2.2.2 *Individual transfer function measurement*

Binaural room impulse responses (BRIRs) and HpTFs were measured and processed for every participant prior to the listening test. Matlab and AKtools[46] were used for signal generation, playback, recording, and processing at a sampling rate of 44.1 kHz. The head positions of the participants were displayed using Pure Data. Communication between the programs was done by UDP messages.

[46] F. Brinkmann and S. Weinzierl (**2017**). "AKtools – An open software toolbox for signal acquisition, processing, and inspection in acoustics" in *142nd AES Convention, Convention e-Brief 309*.

Before starting, participants put on the headphones and were familiarized with the measurement procedure. Their current head position, given by azimuth and x/y/z coordinates was displayed on the LCD screen along with the target azimuth. The head tracker was calibrated with the participant looking at a frontal reference position marked on the LCD screen. Participants were instructed to keep their eye level aligned to the reference position during measurement and listening test, this way establishing also an indirect control over their head elevation and roll. For training proper head-positioning, participants were instructed to move their head to a specific azimuth and hold the position for 10 seconds. A visual inspection showed that all participants were quickly able to maintain a position with a precision of $\pm 0.2^\circ$ azimuth, and ± 2 mm translation in x/y/z coordinates.

Then, participants inserted the ear-molds with measurement microphones into their ear canals until they were flush with the bottom of the concha, and the correct fit was inspected by the investigator. BRIRs were measured for azimuthal head-above-torso orientations within $\pm 34^\circ$ in 2° steps providing a perceptually smooth adaption to head movements[47]. The range allowed for a convenient view of the LCD screen at any head orientation. Sine sweeps of an FFT order 18 were used for measuring transfer functions, with the level of the measurement signal being identical across participants. It was set so to avoid limiting of the DSP-driven loudspeakers, and headphones, and to achieve a peak-to-tail signal-to-noise ratio (SNR) of approx. 80 dB for ipsilateral and 60 dB for contralateral sources without averaging[48]. Because the ear-molds significantly reduced the level at

[47] A. Lindau and S. Weinzierl (**2009**). "On the spatial resolution of virtual acoustic environments for head movements on horizontal, vertical and lateral direction" in *EAA Symposium on Auralization*.

[48] S. Müller and P. Massarani (**2001**). "Transfer function measurement with sweeps. directors cut including previously unreleased material and some corrections" *J. Audio Eng. Soc.* (Original release).

the ear drums, all participants reported it to be still comfortable.

The participants started the measurement by pressing a button on the MIDI interface after moving their head to the target azimuth with a precision of $\pm 0.1^\circ$. For the frontal head orientation, the reference position had to be met within 0.1 cm for the x/y/z-coordinates. For all other head orientations the translational positions naturally deviate from zero; in these cases, participants were instructed to meet the targeted azimuth only, and to move their head in a natural way. During the measurement, head movements of more than 0.5° or 1 cm caused a repetition, which rarely happened. The tolerances were set to avoid audible artifacts introduced by imperfect positioning[49,50]

Thereafter, ten individual HpTFs were measured per participant. Although the headphones were worn during the entire experiment, their position on the participants' head might change due to head movements. To account for corresponding changes in the HpTFs, participants were instructed to rotate their head to the left and right in between measurements. After the measurements, which took about 30 minutes, the investigator carefully removed the in-ear microphones without changing the position of the headphones.

[49] J. Blauert (**1997**). *Spatial Hearing. The psychophysics of human sound localization.*

[50] T. Hiekkanen, et al. (**2009**). "Virtualized listening tests for loudspeakers" *J. Audio Eng. Soc.*

2.2.3 *Post processing*

As a first step, leading zeros in the BRIRs were removed, while the temporal structure remained unchanged. For this purpose, time-of-arrivals (TOAs) were estimated using onset detection, and removed by means of a circular shift. TOA outliers were corrected by fitting a second order polynom or smoothing splines to the TOA estimates—whatever gave the best fit to the valid data (determined by visual inspection). ITDs, i.e. differences between left and right ear TOAs, were re-inserted in real time during the listening test to avoid comb-filter effects occurring in dynamic auralizations with non-time-aligned BRIRs and reducing the overall system latency[51]. In a second step, BRIRs were truncated to 0.4, 1, and 3 seconds for the anechoic, dry and wet environment to allow for a decay of around 60 dB. A squared sine fade out was applied at the intersection between the impulse response decay and the noise floor to artificially extend the decay.

[51] A. Lindau, et al. (**2010**). "Individualization of dynamic binaural synthesis by real time manipulation of the ITD" in *128th AES Convention, Convention Paper*.

Individual HpTF compensation filters of FFT order 12 were designed based on the average HpTF using frequency dependent regularized least mean squares inversion[52]. Regularization was used to limit filter gains if perceptually required: HpTFs typically show distinct notches at high frequencies which are most likely caused by anti-resonances of the pinna cavities[53]. For an example see Fig 2.2 (top) at approx. 10 and 16 kHz. The exact frequency and depth of these notches strongly depends on the current fit of the headphones. Already a slight change in position might considerably detune a notch, potentially leading to ringing artifacts of the applied headphone filters[54]. Therefore, individual regularization functions were composed by manually fitting one to three parametric equalizers (PEQs) per ear to the most disturbing notches. The compen-

[52] S. G. Norcross, et al. (**2006**). "Inverse filtering design using a minimal phase target function from regularization" in *121th AES Convention, Convention Paper 6929*.

[53] H. Takemoto, et al. (**2012**). "Mechanism for generating peaks and notches of head-related transfer functions in the median plane" *J. Acoust. Soc. Am.*

[54] A. Lindau and F. Brinkmann (**2012**). "Perceptual evaluation of headphone compensation in binaural synthesis based on non-individual recordings" *J. Audio Eng. Soc.*

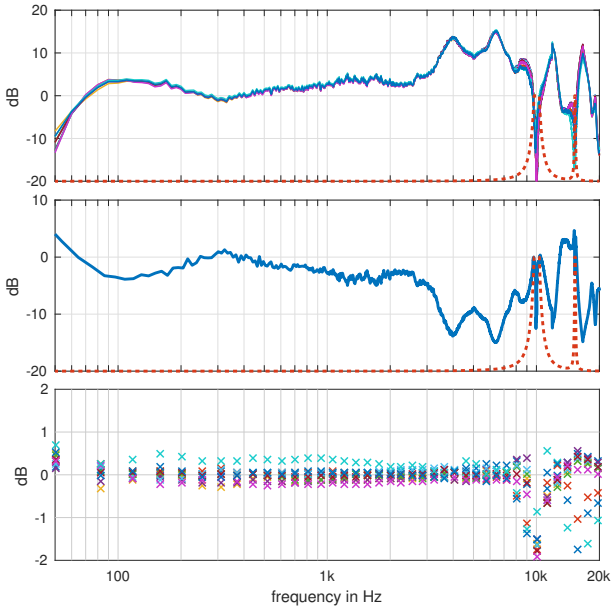


Figure 2.2: Example of the headphone compensation process for the left ear of participant 3. Top: HpTFs (solid lines) and regularization (dashed line). Middle: compensation filter (solid line) and regularization (dashed line). Bottom: difference between compensated HpTFs and target band pass in auditory filters.

sated headphones approached a minimum phase target band-pass consisting of a 4th order Butterworth high-pass with a cut-off frequency of 59 Hz and a second order Butterworth low-pass with a cut-off frequency of 16.4 kHz. The result, i.e. the convolution of each HpTF with the inverse filter, deviated from the target band-pass by less than ± 0.5 dB in almost all cases, except for frequencies where notches in the HpTF occurred (cf. Figure 2.2). The frequency responses of the in-ear microphones remained uncompensated in the BRIRs and HpTFs. This way the inverse frequency responses are present in the HpTF filters, and the microphones influence cancels out if the HpTF filters are convolved with the BRIRs.

Finally, presentations of the real loudspeaker and the binaural simulation had to be matched in loudness. Assuming that signals obtained via individual binaural synthesis closely resemble those obtained from loudspeaker reproduction in the temporal and spectral shape (cf. Figure 2.3), loudness matching can be achieved by simply matching the RMS-level of simulation and real sound field. Hence, 5 s pink noise samples were recorded from loudspeakers and headphones while the participant's head was in the frontal reference position. Before matching the RMS-level, the headphone recordings were convolved with the frontal incidence BRIR and the headphone compensation filter to account for the reproduction paths during the listening test. The loudspeaker recordings were convolved with the target bandpass that was used for designing the headphone compensation filter.

2.2.4 *Test Procedure*

Nine participants with an average age of 30 years (6 male, 3 female) participated in the listening test, all of them experienced with dynamic binaural synthesis. No hearing anomalies were known, and

with a musical background of 13 years on average, all participants were regarded as expert listeners.

The test procedure was identical across the three acoustical environments, and tests were conducted over a period of 20 months. At first, participants were placed on the chair, took on the headphones, and were familiarized with the user interfaces showing the current head position and answer buttons, and the MIDI interface for their control. Afterwards, the BRIRs and HpTFs were measured and processed as described above.

The perceptual testing started with four ABX tests for authenticity per participant (2 sources $\times$ 2 contents) each consisting of 24 trials. The order of content and source was randomized and balanced across participants. At each trial, the binaural simulation and the real sound field were randomly assigned to three buttons (A/B/X, with each condition assigned at least once), and participants started and stopped the audio playback by pressing a button on the MIDI interface. Stopping the playback could also be used to listen to the entire decay in the BRIR. The ABX test is a 3-Interval/2-Alternative Forced Choice (3I/2AFC) test, with the three intervals *A*, *B*, and *X*, and the two possible answers (forced choices) *A equals X*, and *B equals X*.

The participants could take their time at will to repeatedly listen to *A*, *B* and *X* in any order and switch at any time. They were moreover instructed to listen at different azimuthal head-above-torso orientations, to focus on different frequency ranges, and that dynamic cues induced by head movements might also help to distinguish between simulation and reality. Because it was not clear which kind of head movements or positions would be helpful, it was left to the participants to find the best head positions/movements for detecting differences. To avoid a drift in the positioning of the participants during the experiment, they were instructed to keep their head at approx. 0° elevation throughout the test, and to move their head to the reference position given by azimuth and x/y/z coordinates between trials. To ensure this, the participants' head position was monitored by the experimenter, who manually enabled each trial. In addition, head positions were recorded in intervals of 100 ms for post-hoc inspection (cf. Section 2.3.2).

Pulsed pink noise and an anechoic male speech recording (5 s) were used as audio content. Speech was chosen as a familiar 'real-life' stimulus including transient components that were supposed to reveal potential flaws in the temporal structure of the simulation. Noise pulses were believed to best reveal flaws related to the spectral shape. To allow for establishing a stable impression of coloration and decay, a single noise pulse with a length of 0.75 s followed by 1 s silence (anechoic and dry environment), and a length of 1.5 s followed by 2 s silence (wet environment) was played in a loop. Noise bursts were faded in and out with a 20 ms squared sine window. The bandwidth of the stimuli was restricted using a 100 Hz high-pass to eliminate the influence of low frequency background noise on the

binaural transfer functions. Previous studies (cf. studies C, and D in Table 2.3) obtained almost identical detection rates for speech and music, which was confirmed by informal listening prior to the current study. We thus limited ourselves to two types of audio content, in order to allow for more variation in other independent variables (source position, spatial environment).

In a next step, qualitative differences between binaural simulation and real sound field were assessed using the Spatial Audio Quality Inventory (SAQI[55]) as implemented in the WhisPER toolbox[56]. Again, participants could directly compare the two test conditions and take their time at will before giving an answer. Audio playback was started and stopped using two buttons labeled A, and B, behind which the simulation and real sound field were hidden, i.e. participants did not know which button toggled the real sound field. The presentation order of the qualities was randomized to avoid order effects. A list with the names, and descriptions of the perceptual qualities was given to all participants beforehand, and questions could be discussed on site. In addition, attributes and their description were also displayed on the screen to avoid any misunderstandings.

[55] A. Lindau, et al. (**2014**b). "A Spatial Audio Quality Inventory (SAQI)" *Acta Acust. united Ac.*

[56] S. Ciba, et al. (**2014**). *WhisPER. A MATLAB toolbox for performing quantitative and qualitative listening tests* `https://dx.doi.org/10.14279/depositonce-31.2`.

The test took about two hours including breaks during which the participants had to remain seated to avoid any change in the test environment that might have introduced additional errors. 30 minutes were needed for binaural measurements, 10 minutes for post processing. The participants took on average 50 minutes for AFC testing ($SD = 11$ minutes), and 21 minutes for the SAQI ratings ($SD = 9$ minutes). The test duration was perceived as just about tolerable by the participants.

Dynamic auralization was realized using the fast convolution engine *fWonder*[57] in conjunction with an algorithm for real-time reinsertion of the ITD[58]. *fWonder* was also used for applying (a) the HpTF compensation filter and (b) the loudspeaker target bandpass. The playback level for the listening test was set to 60 dB SPL_{AFeq}. This way the 60 dB dynamic range in the measured BRIRs ensures that their decay continues approximately until the absolute threshold of perception around 0 dB SPL is reached. Any artifacts related to the truncation of the measured BRIRs were thus expected to be inaudible. BRIRs used in the convolution process were dynamically exchanged according to the participants' current azimuthal head-above-torso orientation (head azimuth), and playback was automatically muted if the participant's head orientation exceeded 35° azimuth.

[57] A. Lindau, et al. (**2007**). "Binaural resynthesis for comparative studies of acoustical environments" in *122th AES Convention, Convention Paper 7032*.

[58] A. Lindau, et al. (**2010**). "Individualization of dynamic binaural synthesis by real time manipulation of the ITD" in *128th AES Convention, Convention Paper*.

2.2.5 *Alternative forced choice test design*

The M-I/N-AFC method provides an objective, criterion-free, and particularly sensitive test for the detection of small differences[59], and thus seems appropriate for testing the authenticity of virtual environments. As a Bernoulli experiment with a guessing rate of 1/N, the binomial distribution allows to calculate the probability that a cer-

[59] L. Leventhal (**1986**). "Type 1 and type 2 errors in the statistical analysis of listening tests" *J. Audio Eng. Soc.*

tain number of correct answers occurs by chance, thus enabling tests on statistical significance: If the amount of correct answers is significantly above chance level, the simulation would *not* be considered as perceptually authentic.

If N-AFC tests are used in the context of authenticity, one should be aware that this corresponds to proving the null hypothesis H_0, i.e., proving that simulation and reality are indistinguishable. Strictly speaking, this proof cannot be given by inferential statistics. The approach commonly pursued is to establish empirical evidence that supports the H_0 by rejecting a minimum effect alternative hypothesis H_1 representing an effect of irrelevant size, i.e. a tolerable increase of the detection rate above the guessing rate[60].

[60] K. R. Murphy and B. Myors (**1999**). "Testing the hypothesis that treatments have negligible effects: Minimum-effect tests in the general linear model" *J. of Applied Psychology*.

The test procedure is usually designed to achieve small type 1 error levels (wrongly concluding that there was an audible difference although there was none) of typically 0.05, making it difficult—especially for smaller differences—to produce significant test results. If we aim, however, at proving the H_0 such a design may unfairly favor our implicit interest ('progressive testing'), that is reflected in the type 2 error (wrongly concluding that there was no audible difference although indeed there was one).

Therefore, we first specified a practically meaningful detection rate of $p_\mathrm{d} = 0.9$ to be rejected, and then aimed at balancing type 1 and type 2 error levels in order to statistically substantiate the rejection *and* the acceptance of the null hypothesis, i.e. the conclusion of authenticity. According to these considerations, a 3I/2AFC listening test design with 24 trials for each participant and test condition was chosen. This lead to a critical value of 18 (75%) or more correct answers in order to reject the $H_0(p_\mathrm{2AFC} = 0.5)$, while for less than 18 correct answers, the specific $H_1(p_\mathrm{2AFC} = 0.9)$ could be rejected (p_NAFC: N-AFC detection rate). Type 1 and type 2 error levels were initially set to 5% and corrected for multiple testing of 4 test conditions by means of Bonferroni correction.

The detection rate of $p_\mathrm{2AFC} = 0.9$ may seem high at first glance, but it corresponds to the expectation that even small differences would lead to high detection rates, considering trained participants and a sensitive test procedure (cf. Leventhal (1986)[61], p. 447) that included suitable audio contents and unlimited listening. Moreover, the critical value of 18 corresponds to the threshold of perception where a participant would identify existing differences in 50% of the cases, which seems to be an adequate criterion for deciding whether or not a simulation is authentic. Note that N-AFC detection rates can be corrected for guessing by $[p_\mathrm{NAFC} - 1/N] \cdot [1/(1 - 1/N)]$.

[61] L. Leventhal (**1986**). "Type 1 and type 2 errors in the statistical analysis of listening tests" *J. Audio Eng. Soc.*

2.2.6 *Qualitative test design*

The German version of the Spatial Audio Quality Inventory (SAQI) was used for assessing detailed qualitative judgements. It consists of 48 perceptual attributes for the evaluation of virtual acoustic environments which were elicited in an expert focus group for virtual

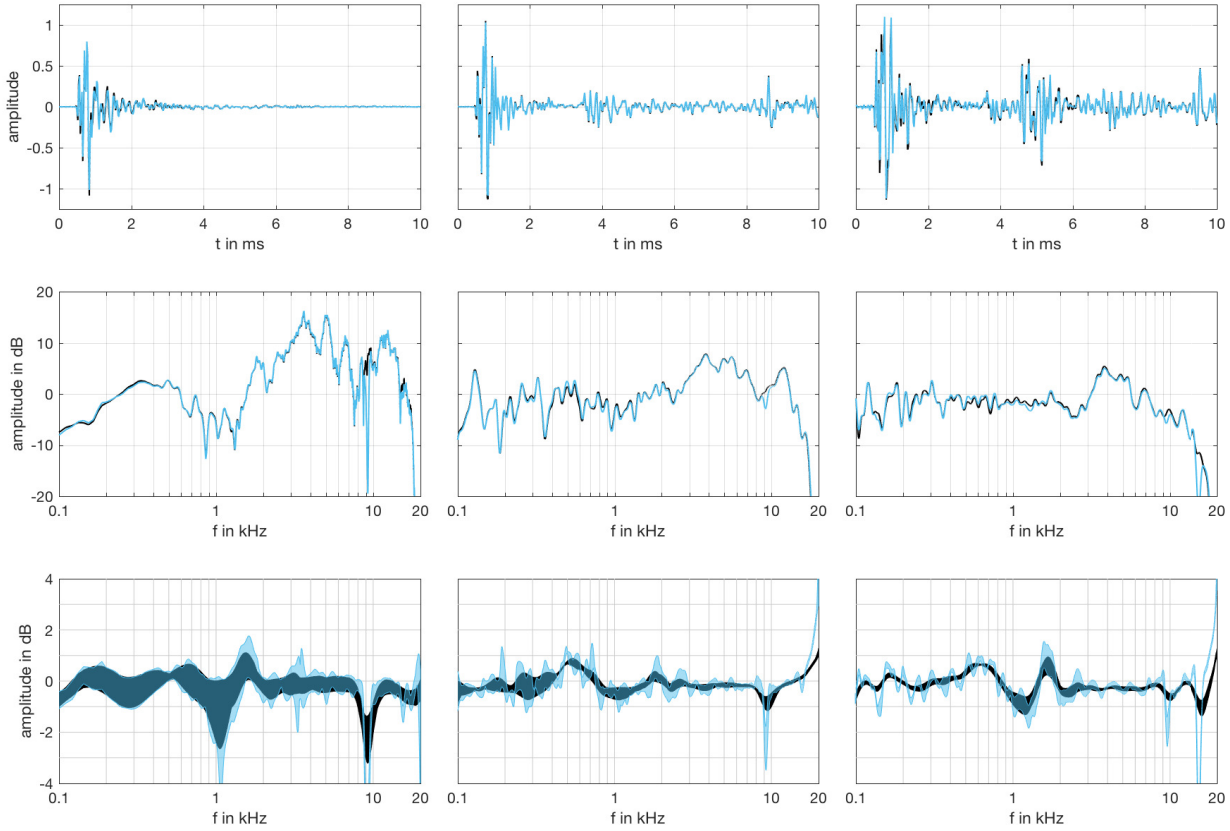


Figure 2.3: Differences between binaural simulation and real sound field for the frontal source measured in the anechoic (left), dry (middle), and wet (right) acoustic environment. Top row shows real (black lines) and simulated (blue lines) binaural impulse responses, middle row real and simulated binaural magnitude spectra for a neutral head-above-torso orientation (12th octave smoothed spectra are shown in case of the dry and wet environment to improve readability). Bottom row shows the range of differences between 12^{th} (light blue) and 3^{rd} octave (dark blue) smoothed magnitude spectra for all head-above-torso orientations.

acoustic environments. Each SAQI quality is accompanied by a short verbal description as well as suitable scale end labels.

We used the SAQI for a direct comparison, i.e. participants rated differences between the simulation and the real sound field, with a rating of zero indicating no perceivable difference. As we were interested in a broad and explorative evaluation, only three qualities of the complete SAQI were excluded, because they were considered irrelevant in our case (*Speed*, *Sequence of events*, *Speech intelligibility*). To limit the time of the listening test to a practical amount, the qualitative evaluation was only carried out for the frontal sound source and the pulsed pink noise.

2.3 *Results*

2.3.1 *Physical evaluation*

Prior to the perceptual evaluation, acoustic differences between the test conditions were estimated based on measurements with the FABIAN dummy head that is equipped with a computer controllable neck joint[62]. Therefore, FABIAN was placed on the chair to measure BRIRs and HpTFs as described in Section 2.2. In a second step, BRIRs were measured as being reproduced by the headphones and the simulation engine as described above. Differences between simulation and real sound field for the left ear and the frontal source are shown

[62] A. Lindau, et al. (**2007**). "Binaural resynthesis for comparative studies of acoustical environments" in *122th AES Convention, Convention Paper 7032*.

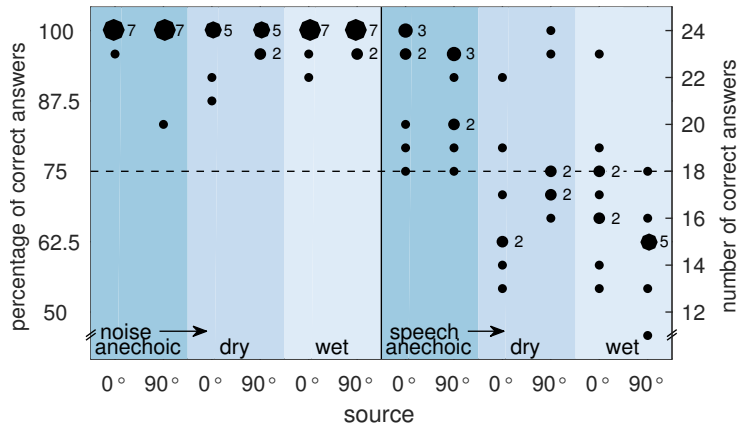


Figure 2.4: Detection rates of the 2AFC test for all participants and test conditions. The size of the dots and the numbers next to them indicates how many participants scored identical results. Results on or above the dashed line are significantly above chance, indicating that differences between simulated and real sound field were audible

in Figure 2.3. They are comparable to the right ear differences, and the lateral source.

Simulated and real BRIRs for the neutral head-above-torso orientation (top row), show a striking similarity for all test environments. For ease of display only the first 10 ms are shown. Corresponding magnitude spectra (middle row) are very similar for the anechoic environment. Slightly higher deviations occur for the reverberant environments in certain frequency ranges (e.g. around 1 kHz), presumably caused by differences in the late part of the BRIRs.

For a better overview, the range of errors for all head-above-torso orientations between $\pm 34°$ is illustrated in the bottom row of Figure 2.3 for 3rd and 12th octave smoothed magnitude spectra. For most frequencies and head orientations, differences are in the range of ± 1 dB which is in good accordance to results of earlier studies[63,64,65,66]. Larger deviations occur at frequencies of about 9 kHz and 16 kHz where narrow and deep notches in the HpTF remained uncompensated for robustness against headphone re-positioning (c.f. Section 2.2.3). However, they exhibit widths of 9% or less relative to their center frequency, and were thus expected to be inaudible; Moore *et al.*[67] reported an audibility threshold of 12.5% relative notch width. Spectral differences are slightly larger in the anechoic environment, in particular at about 1 kHz. At this frequency a notch appears for the left ear in case the head is turned away from the source—i.e. for head-above-torso orientations in the range of 30°. This notch originates from delayed copies of the sound traveling around the head on different paths[68].

Assuming that third octave differences in the range of 0.5 dB might already be audible for expert listeners and sensitive listening test designs (compare ΔG_{95} in Table 3 from[69]), we can expect that the binaural simulation will turn out to be not perceptually authentic, at least for the noise content[70].

2.3.2 *Perceptual authenticity*

The detection rates of the 2AFC test are summarized in Figure 2.4 for all participants and test conditions. Although statistical analysis

[63] E. H. A. Langendijk and A. W. Bronkhorst (**2000**). "Fidelity of three-dimensional-sound reproduction using a virtual auditory display" *J. Acoust. Soc. Am.*

[64] A. H. Moore, et al. (**2010**). "An initial validation of individualised crosstalk cancellation filters for binaural perceptual experiments" *J. Audio Eng. Soc.*

[65] F. L. Wightman and D. J. Kistler (**1989**a). "Headphone simulation of free field listening. I: Stimulus synthesis" *J. Acoust. Soc. Am.*

[66] D. Pralong and S. Carlile (**1996**). "The role of individualized headphone calibration for the generation of high fidelity virtual auditory space" *J. Acoust. Soc. Am.*

[67] B. C. J. Moore, et al. (**1989**). "Detection and discrimination of spectral peaks and notches at 1 and 8 kHz" *J. Acoust. Soc. Am.*

[68] V. R. Algazi, et al. (**2001**a). "Elevation localization and head-related transfer function analysis at low frequencies" *J. Acoust. Soc. Am.*

[69] F. Brinkmann, et al. (**2015**b). "Audibility and interpolation of head-above-torso orientation in binaural technology" *IEEE J. Sel. Topics Signal Process.*

[70] See Supplementary materials at `http://dx.doi.org/10.14279/depositonce-8510` for example auralizations.

		anechoic		dry		wet		all
		0°	90°	0°	90°	0°	90°	
A	**noise**	99.5	97.9	97.0	98.8	98.6	99.1	98.5
	speech	91.2	87.5	68.5	79.1	71.3	61.6	76.2
B	**noise**	10.3	9.1	15.2	15.1	12.3	13.2	12.4
	speech	30.3	29.6	43.4	39.1	35.9	40.7	36.3
C	**noise**	6.6	5.6	8.5	9.1	10.6	12.6	8.9
	speech	21.6	16.6	18.6	28.2	27.3	19.0	21.8

Table 2.2: Data from the 2AFC listening test averaged across participants. (A) 2AFC detection rates in percent. (B) Rating duration per trial in seconds (additionally averaged across trials). (C) Amount of head movements specified by the difference $P_{75} - P_{25}$ of observed head azimuths in degree (P_i: i_{th} percentile).

of authenticity was conducted on the level of individual participants, the observed average detection rates are given in Table 2.2 A for better comparability to earlier studies, and because the corresponding detection frequencies were used to statistically analyze effects between test conditions by means of χ^2 tests. One participant could not participate in the anechoic environment due to illness, and two participants who accidentally touched the headphones after the binaural measurements were excluded from the results of the dry environment. Both reported hissing sounds that might be attributed to ringing artifacts caused by headphone repositioning.

A clear difference in detection performance was found between the audio contents: For pulsed noise, all participants were able to discriminate simulation and real sound field, i.e. all individual detection rates are above the dashed line in Figure 2.4. For the speech stimulus, however, the simulation turned out to be authentic in 44% of cases (dots below dashed line). χ^2 tests showed this effect to be statistically highly significant ($\chi^2 = 32.81$, $p < 0.001$, $df = 1$). A significant effect of the room was observed in interaction with the speech content, where all participants detected differences in the anechoic environment, whereas only 43%, and 28% detected differences in the dry and the wet environment, respectively ($\chi^2 = 8.46$, $p = 0.001$, $df = 2$). Pairwise comparisons showed significant differences between the anechoic and wet room ($\chi^2 = 7.96$, $p = 0.005$, $df = 1$) and almost significant differences between the anechoic and dry room ($\chi^2 = 3.57$, $p = 0.059$, $df = 1$). Differences between the dry and wet room were statistically insignificant with the given sample size and test power ($\chi^2 = 0.88$, $p = 0.349$, $df = 1$). In line with earlier studies, no significant effect was found for the source position ($\chi^2 = 0.04$, $p = 0.84$, $df = 1$).

Differences between audio contents were also found in the rating durations and the amount of azimuthal head movements. An inspection reveals that giving an answer for a trial took the participants about three times longer when listening to the speech content (Table 2.2 B), and that they used larger head movements to detect differences (Table 2.2 C). Both effects are highly significant in all environments (Bonferroni corrected Wilcoxon signed rank tests for dependent samples, $p \leq 0.01$, $\alpha = 0.05$). This is also reflected by the high and significant correlations between the average detection rates (Table 2.2 A) and i) the rating duration (Pearson correla-

tion, $r = -0.92$, $p < 0.01$), and ii) the amount of head movements ($r = -0.73$, $p < 0.01$): Participants with lower detection rates took more time and moved their head further when trying to detect differences between reality and simulation. Participants who could not reliably detect differences, on average explored 95% of the available range of head-above-torso orientations. Thus it is unlikely that the results are biased due to an insufficient exploration of the binaural simulation. Again no effects for the source position were observed, whereas effects of the acoustic environment are rather small but significant at least for the rating duration between the anechoic and dry, as well as the anechoic and wet room (Bonferroni corrected Wilcoxon signed rank tests for dependent samples, $p \leq 0.05$, $\alpha = 0.05$).

During auralization, BRIRs were selected solely by the participants' head azimuth. Hence, unobserved differences with respect to the measured positions in the remaining degrees of freedom—translation in x, y, z, elevation, lateral flexion—might have caused audible artifacts. Therefore, the recorded head positions of all participants were used for a post hoc analysis of deviations between head position during binaural measurements and 2AFC tests. For the translation in x, y, and z coordinates, deviations were found to be smaller than 1 cm for about 95% of the time and never exceeded 2 cm. Differences in head elevation (tilt) and in lateral flection (roll) rarely exceeded 10° and were below 5° for 90% of the time. While this may have caused audible artifacts occasionally, a systematic influence of the results is highly unlikely[71],[72].

[71] J. Blauert (**1997**). *Spatial Hearing. The psychophysics of human sound localization* (p. 44).

[72] T. Hiekkanen, et al. (**2009**). "Virtualized listening tests for loudspeakers" *J. Audio Eng. Soc.*

2.3.3 *Qualitative evaluation*

The qualitative analysis with respect to 45 perceptual attributes is summarized in Figure 2.5(a). They were only assessed for the frontal source and the pulsed pink noise to limit the duration of the listening test. Please note that the scale labels (y-labels) were omitted for better readability. They can be found in Table I in Lindau *et al.*[73], whereby a rating of -1 refers to the first, and a rating of 1 to the second label. A rating of 0 indicates no perceptual difference between simulation and real sound field. Because the ratings were not normally distributed in 60% of the cases (Shapiro-Wilk tests, $p \leq 0.2$), Figure 2.5 shows the median values, interquartile ranges (IQR), and the total range.

[73] A. Lindau, et al. (**2014**b). "A Spatial Audio Quality Inventory (SAQI)" *Acta Acust. united Ac.*

In line with the results of the AFC test, several findings indicate that the simulation has a high degree of realism with respect to almost all tested perceptual aspects: (i) The IQR does include zero in almost all cases (ii) The median is zero in 92% of the cases, and (iii) all participants made zero ratings for seven qualities (*roughness*, *doppler effect*, *front/back position*, *pre-echos*, *noise-like artifacts*, *alien source*, *distortion*), while 14 more qualities obtained zero ratings from all participants in at least one environment (*metallic tone color*, *level of reverberation*, *duration of reverberation*, *envelopment*, *spatial disintegration*, *post-echos*, *temporal disintegration*, *responsiveness*, *dynamic range*, *compression*, *pitched artifact*, *impulsive artifact*, *ghost source*, *tactile vibration*).

Larger differences (IQRs that don't overlap zero) were found for the *difference*, which confirms the results of the 2AFC test, the *tone color bright/dark*, where the negative ratings indicate that the simulation was perceived to be darker than the real sound field in the anechoic environment, and the *horizontal direction* in the dry room. Apart from the latter two cases, IQRs overlap each other for all test conditions, suggesting that differences between rooms are rather small.

In tendency, the participants' ratings indicate that the simulation has slightly less *naturalness*, *clarity*, and *presence*, and that the real sound field was preferred over the simulation (*liking*). However, median values were zero for the above mentioned qualities, except for *liking* in the anechoic environment. Apart from that, non-zero median values, or large IQRs were only found in the categories *tone color*, *tonalness*, and *geometry*, in turn suggesting that there are no relevant deficits regarding *room*, *time*, *dynamics*, or *artifacts* of any kind.

In some cases, equally distributed positive and negative ratings could conceal perceptually relevant differences if they result in a zero median. To uncover this effect, distributions for all absolute ratings with median values ≥ 0.05 are shown in Figure 2.5(b)—sorted in descending order to emphasize their relevance. Besides the overall *difference*, three perceptual qualities related to coloration (*high frequency tone color*, *tone color bright-dark*, and *pitch*), as well as *distance* show systematic deviations from zero. However, the IQRs already include zero for *pitch* and *distance*.

2.4 Discussion

At least four empirical studies were concerned with the authenticity of binaural simulations, i.e. with the physical and perceptual identity of ear signals produced by natural acoustic environments and their equivalent produced by binaural synthesis: Langendijk and Bronkhorst[74] (termed *A* in the following), Moore *et al.*[75] (*B*), Masiero[76] (*C*), and Oberem *et al.*[77] (*D*). In contrast to all previous studies, which used static synthesis in anechoic conditions, the current investigation considered, for the first time, dynamic binaural synthesis, allowing for natural head movements of the listeners, as well as a sample of three different acoustic environments with different degrees of reverberation. The results can thus be expected to have more ecological validity with respect to the large variety of current and future applications of binaural technology.

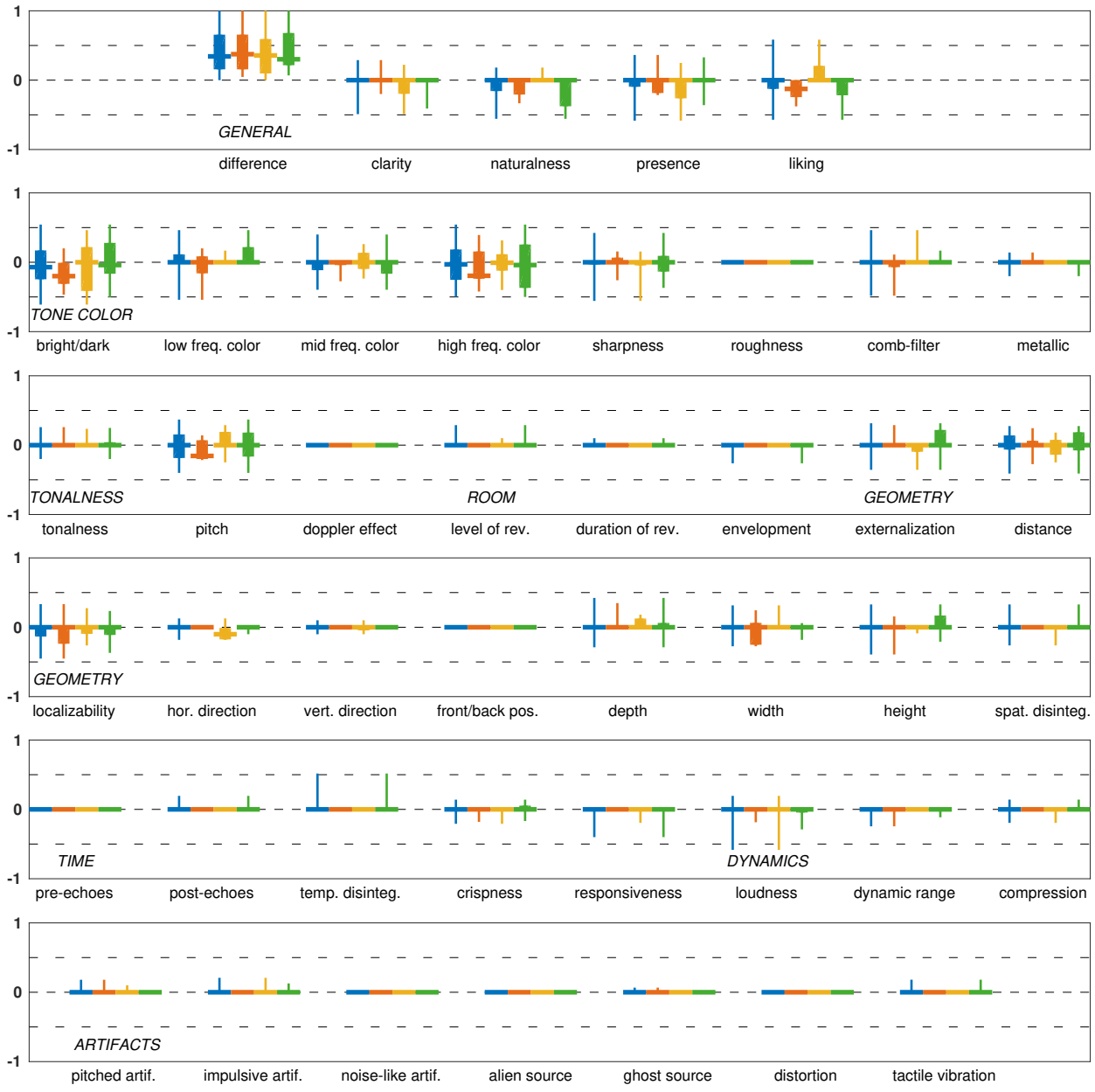
The physical identity, i.e. the extent to which inaccuracies of the binaural reconstruction could be controlled, was similar in all studies. In terms of magnitude deviations between real and simulated binaural transfer functions, comparable values have been reported: *A* found 12^{th} octave magnitude differences of ± 1 dB for ipsilateral sources and ± 5 dB for contralateral sources, along with phase differences of up to 6°. *B* reported magnitude deviations in the unsmoothed spectra to be smaller than ± 2 dB except for frequencies above 6 kHz where deep notches occurred. We found 12^{th} octave

[74] E. H. A. Langendijk and A. W. Bronkhorst (**2000**). "Fidelity of three-dimensional-sound reproduction using a virtual auditory display" *J. Acoust. Soc. Am.*

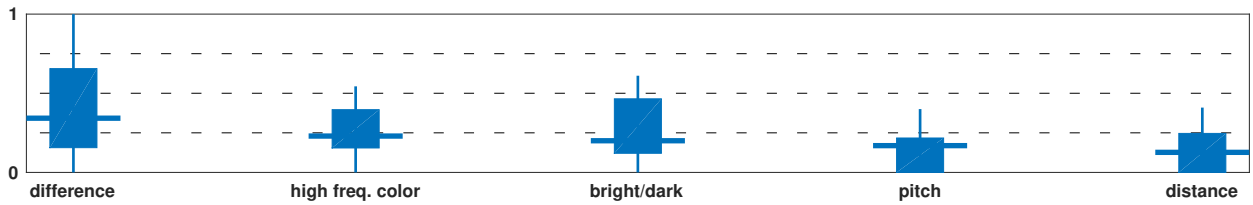
[75] A. H. Moore, et al. (**2010**). "An initial validation of individualised crosstalk cancellation filters for binaural perceptual experiments" *J. Audio Eng. Soc.*

[76] B. Masiero (**2012**). "Individualized binaural technology. measurement, equalization and perceptual evaluation" Doctoral Thesis.

[77] J. Oberem, et al. (**2016**). "Experiments on authenticity and plausibility of binaural reproduction via headphones employing different recording methods" *Appl. Acoust.*



(a)



(b)

Figure 2.5: SAQI ratings by means of median (horizontal lines), IQR (boxes), and overall range (vertical lines). (a) The four bars for each perceptual quality show results pooled across rooms, and for the anechoic, dry and wet room (from left to right). (b) Pooled absolute ratings with median values ≥ 0.05 in descending order.

	Langendijk (A)	Moore (B)	Masiero (C)[a]	current study
Detection rate[b]	53.3% noise	59.4/59.4/48% noise/pulses/tones	87.5/71.4/73.7% noise/speech/music	see Table 2.2
Critical det. rate[c]	52.33% (1800 trials, 1 test)	59.38% (192 trials, 6 tests)	55.4/55.19/55.65% (208-241 trials, 3 tests)	assessed on participant basis
Audio content	Noise, 0.5-16 kHz, varying spectral shape	Noise, 0.12-15 kHz Pulse trains, 0.1-15 kHz Complex tone, 0.1-4.6 kHz	Noise, 0.2-20 kHz Speech, 0.2-8 kHz Music, 0.2-10 kHz	Noise, 0.1-16.4 kHz, Speech, 0.1-16.4 kHz
Test Environment	static synthesis extraaural headphones anechoic 6 sources, (around listener)	static synthesis CTC loudspeakers anechoic 1 source, (frontal)	static synthesis circumaural headphones anechoic 24 sources, (around listener)	dynamic synthesis extraaural headphones anechoic & reverberant 2 sources, (frontal & lateral)
Test method	4I/2AFC listening once with training, with feedback 6 participants, (experienced)	4I/2AFC listening once with training with feedback 8 participants, (mostly experienced)	3I/3AFC listening three times without training without feedback 40 participants, (unexperienced)	3I/2AFC unlimited listening with training without feedback 9 participants, (experienced)

[a]Detection rates from *D* (experiment with blocked ear canal measurements): 79.3% (noise), 66.8% (speech), 69.8% (music).
Critical detection rate: 52.75% (800 trials, 3 tests).

[b]Averaged across participants and sources; detection rates from *C*, and *D* were transformed to 2AFC detection rates.

[c]Dunn-Šidák correction for multiple testing was applied to the initial type 1 error level of 5%.

Table 2.3: Overview of studies on perceptual authenticity of binaural synthesis. Sorted by ascending detection rates from left to right, and named according to first author. For the current study, only the results for the anechoic environment are listed. See text for details.

magnitude differences to be smaller than ± 1 dB for most frequencies (cf. Figure 2.3). Deviations of comparable magnitude were also observed in studies that focused *only* on the physical accuracy in the reproduction of binaural signals[78,79,80]. We can thus conclude that this seems to be the degree to which the physical identity of reality and simulation can be brought by carefully controlling the measurement, processing, and reproduction of binaural simulations.

The perceptual identity of acoustic environments and their binaural simulations could not be observed, neither in previous nor in the current study. To investigate this, all studies used M-Intervall, N-Alternative Forced Choice tests (M-I/N-AFC). An overview of the results is shown in Table 2.3, with results from *D* given as a footnote, because the test was conducted in the same laboratory as *C* with identical audio content, test environment, and test method. Since studies *C* and *D* did not conduct significance tests, the critical detection rates were computed based on the reported test method, and transformed to 2AFC rates afterwards. The overview shows that: i) the detection rates were significantly above chance level except for a synthetic and strongly band limited complex tone in *B*, ii) the equivalent 2AFC detection rates for noise span from 52% in *A* (2% above chance level) to 87.5% in *B*, and iii) detection rates of previous studies were generally lower than those observed in the current study.

Since the physical accuracy was comparable in all studies, there are three factors which can account for the considerable differences in the measured detection rates. This is: i) the presented audio con-

[78] F. L. Wightman and D. J. Kistler (**1989**a). "Headphone simulation of free field listening. I: Stimulus synthesis" *J. Acoust. Soc. Am.*

[79] C. Ryan and D. Furlong (**1995**). "Effects of headphone placement on headphone equalisation for binaural reproduction" in *98th AES Convention*.

[80] D. Pralong and S. Carlile (**1996**). "The role of individualized headphone calibration for the generation of high fidelity virtual auditory space" *J. Acoust. Soc. Am.*

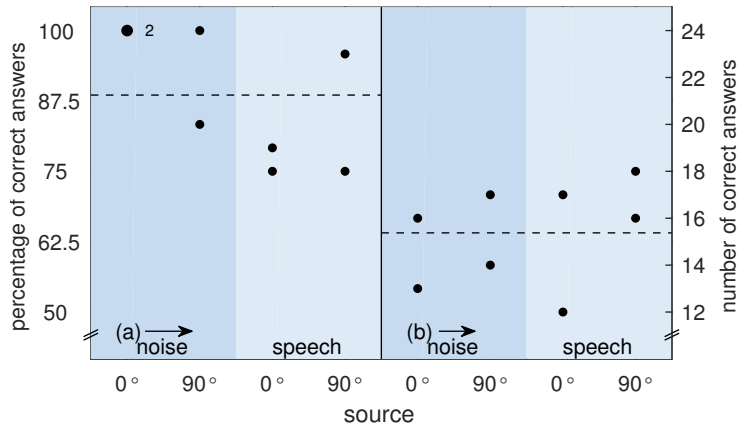


Figure 2.6: The 2AFC detection rates for two participants and the anechoic environment. (a) Results for dynamic binaural synthesis and listening at will; (b) results for static synthesis and listening only once. The size of the dots and the numbers next to them indicates how many participants scored identical results. Dashed lines show averaged detection rates

tent, ii) the technical implementation of the listening test, and iii) the exact test procedure of listening and decision making.

With respect to the audio content, stimuli with a broadband, steady, and non-sparse spectrum (such as noise or pulses) produced considerably and significantly higher detection rates than music or speech. This was shown by C, with 87.5% for noise compared to 71.4% for speech and 73.7% for music, as well as by the current study, with 98.5% for noise and 76.2% for speech. This can be attributed to a surplus of physical cues facilitating the identification of small timbral differences, which were shown to be mainly responsible for the detection performance by the qualitative evaluation. Obviously, these spectral cues were outweighed neither by the more transient character of speech, nor by the higher familiarity with speech as an everyday stimulus, since the listener does not have to draw on his or her internal reference and experience in an N-AFC listening test providing an immediate comparison between simulation and reality.

Comparing the technical setup of studies *A*, *B*, *C*, and *D*, two differences become apparent. First, the circumaural headphones used by *C*, and *D* were the only ones that had to be repositioned after measuring the binaural signals. Since even naïve listeners are able to reliably detect differences due to headphone repositioning[81], it is almost certain that this effect considerably increased the resulting detection rates. In addition, participants in *D* were re-seated between binaural measurements and listening test, which is also likely to introduce audible artifacts[82]. To avoid this, we used extraaural headphones that remained in position during the entire listening test, and allowed for removing the in-ear microphones without moving the headphones. Moreover, participants remained in position during the entire listening test, and their position was monitored with a high precision head tracking system.

With respect to the test method, participants could listen only once to a sequence of stimuli before giving an answer in *A* and *B*, compared to listening up to three times in *C* and *D*, or unlimited listening as allowed in our study, where the subjects could listen as often as they wanted, in any order, and switch between stimuli at any time. To investigate the extent to which this difference increased

[81] M. Paquier and V. Koehl (**2015**). "Discriminability of the placement of supraaural and circumaural headphones" *Appl. Acoust.*

[82] T. Hiekkanen, et al. (**2009**). "Virtualized listening tests for loudspeakers" *J. Audio Eng. Soc.*

the sensitivity of the test, we conducted a 3I/2AFC test in the anechoic environment and sequentially presented the stimuli only once. Two participants, who also took part in the previous test a couple of days earlier, were selected to allow for a direct comparison of their detection rates under both conditions. As shown in Figure 2.6, the changed mode of presentation caused a considerable and statistically significant decrease in detection rates from 89% for unrestricted listening to 64% for restricted listening (single sided Wilcoxon signed rank test for dependant samples, paired across audio contents and source positions, $p < 0.01$). The detection rates were now similar for noise (62.5%) and speech (65.6%), and they are well comparable to the detection rates of earlier studies. These results emphasize the impact of the test design and point out that a direct comparison of detection rates across studies has to be carried out with caution.

Interestingly, neither *A*, nor *C*, and *D* reported notable effects of the source position. In the current study, listeners could change their head orientation and were thus not restricted to a fixed relative source position anyway. Also in this case, the two source positions provided (frontal/lateral) did not entail significantly different detection rates.

2.5 *Conclusion*

In the present study, we assessed whether the binaural re-synthesis of electro acoustic sources can be discriminated from the corresponding real sound field under optimal test conditions, such as individual BRIRs and no headphone repositioning, as well as state-of-the-art measurement, post-processing, and rendering. For the first time, perceptual 'authenticity' was tested for spatial environments with different degrees of reverberation, and dynamic binaural simulations accounting for azimuthal head-movements of the listeners. Remaining differences were evaluated with respect to the acoustical signal as well as with a finely differentiated inventory of perceptual attributes.

For testing the perceptual authenticity at the level of individual participants, we conducted a minimum effect size N-AFC listening test with balanced type one and type two errors. In order to maximize the sensitivity of the test, we allowed for repeated listening and switching between the stimuli, and provided audio content suitable to uncover different potential flaws of the simulation. The influence of the acoustical environment (anechoic/dry/wet) and the source position (frontal/lateral) were analyzed as independent variables.

In agreement with earlier studies, we found that—even with these prerequisites—for a pulsed pink noise sample all participants could reliably detect differences between reality and simulation. For the speech sample, however, the detection rates of individual participants ranged from 62% to 91% with a mean of 76%. Hence, for almost half of the trained expert listeners, who could immediately and repeatedly compare the two stimuli under optimal listening conditions, the simulation can be considered as perceptually authentic.

An interaction between audio content and room was observed, with detection rates being highest for the anechoic environment and lowest for the reverberation chamber in case of the speech content. The remaining differences between simulation and reality, that stem from measurement uncertainties, manifest themselves mainly in a degradation of tone color related qualities rather than in localization or spatial impression. This in turn explains why pink noise with its broadband spectral content provides the strongest cues to identify these differences.

In both analyses (technical and the perceptual), the anechoic condition proved to be the worst case for binaural re-synthesis. With increasing reverberation time and constant source-receiver distances, i.e. with decreasing direct-to-reverberant energy ratios (DRRs), the spectral differences between reality and simulation are at least partially smoothed out, corresponding to lower detection rates for the reverberant environments compared to the anechoic situation. The difference between the frontal and lateral source position, on the other hand, had no significant influence for any of the three spatial environments.

The results suggest that for 'everyday audio content' with limited spectral bandwidth such as speech and music, an authentic virtual representation of acoustic environments can be achieved by using individual dynamic binaural synthesis, if sufficient care is taken for the acquisition, post-processing and rendering of the corresponding binaural impulse response datasets. While natural acoustic sources with their time-variant behavior present particular challenges, this should always be possible for loudspeakers and electro-acoustic reproduction systems and enable their perceptual evaluation by binaural re-synthesis without a relevant loss in quality.

3
Audibility and interpolation of head-above-torso orientation in binaural technology

Fabian Brinkmann, Reinhild Roden, Alexander Lindau, and Stefan Weinzierl (**2017**), *IEEE J. Sel. Topics Signal Process.*, **9**(5), 931–942. DOI: 10.1109/JSTSP.2015.2414905.
(Accepted manuscript. ©2015 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.)

Head-related transfer functions (HRTFs) incorporate fundamental cues required for human spatial hearing and are often applied to auralize results obtained from room acoustic simulations. HRTFs are typically available for various directions of sound incidence and a fixed head-above-torso orientation (HATO). If – in interactive auralizations – HRTFs are exchanged according to the head rotations of a listener, the auralization result most often corresponds to a listener turning head *and* torso simultaneously, while – in reality – listeners usually turn their head *independently* above a fixed torso. In the present study, we show that accounting for HATO produces clearly audible differences, thereby suggesting the relevance of correct HATO when aiming at perceptually transparent binaural synthesis. Furthermore, we addressed the efficient representation of variable HATO in interactive acoustic simulations using spatial interpolation. Hereby, we evaluated two different approaches: interpolating between HRTFs with identical torso-to-source but different head-to-source orientations (*head interpolation*) and interpolating between HRTFs with the same head-to-source but different torso-to-source orientations (*torso interpolation*). Torso interpolation turned out to be more robust against increasing interpolation step width. In this case the median threshold of audibility for the head-above-torso resolution was about 25 degrees, whereas with head interpolation the threshold was about 10 degrees. Additionally, we tested a non-

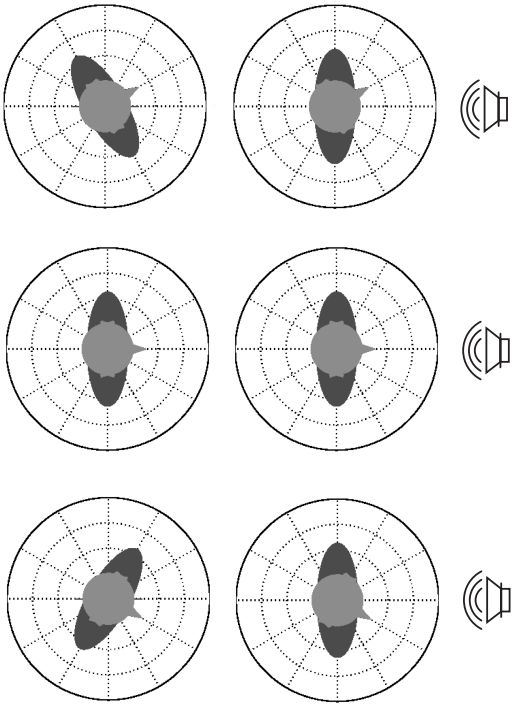


Figure 3.1: Illustration of head rotations with constant (left) and variable (right) HATO and head orientations of 30° (top), 0° (middle), and 330° (bottom). HATO is always 0° for the head rotation displayed in left column and otherwise equals the displayed head orientation.

interpolation approach (nearest neighbor) as a suitable means for mobile applications with limited computational capacities.

3.1 Introduction

Interactive auralization, such as dynamic binaural synthesis, accounts for head rotations of the listener by real-time exchange of corresponding binaural transfer functions. Rendering is often based on sound fields obtained from room acoustic simulations, making it possible to auralize rooms while using arbitrary HRTF sets. Binaural room impulse responses (BRIRs) required for auralization are then obtained by superposition of head-related impulse responses (HRIRs) corresponding to the respective incident angles of direct sound and reflections (Vorländer, 2008, p. 272[1]). Interactivity with respect to head rotations fosters a realistic overall impression, helps in resolving front-back confusions [2], and when judging timbre [3]. However, HRTFs usually represent different angles of sound incidence relative to a *fixed* dummy head or human subject. At the reproduction stage, head rotations will thus correspond to a listener moving head *and* torso whereas in a typical situation the head is rotated *independently* above a fixed torso (Figure 3.1).

The effect of the torso on HRTFs was extensively studied by Algazi et al. [4] for static binaural synthesis and a neutral head-above-torso orientation (HATO). The authors showed that if the torso blocks the direct path from the sound source to the ear, shadowing occurs for frequencies above approximately 100 Hz, causing increasing attenuation of up to 25 dB. For other directions of sound incidence, the

[1] M. Vorländer (**2008**). *Auralization. Fundamentals of acoustics, modelling, simulation, algorithms and acoustic virtual reality*.

[2] P. Minnaar, et al. (**2001**). "The importance of head movements for binaural room synthesis" in *International Conference on Auditory Display*.

[3] J. Blauert (**1997**). *Spatial Hearing. The psychophysics of human sound localization*.

[4] V. R. Algazi, et al. (**2002**). "Approximating the head-related transfer function using simple geometric models of the head and torso" *J. Acoust. Soc. Am.*

torso acts as a reflector causing comb-filters with an amplitude of up to ± 5 dB, whereas the exact positions of peaks and dips of the comb-filter mainly depends on the source elevation. For a source above the listener the first dip occurs already at a frequency as low as 700 Hz. While the torso influence can be shown to extend across the complete audio range, pinnae cues increasingly dominate the spectral shape of the HRTF above 3 kHz (variations up to approx. ± 20 dB) [5,6].

From an analysis of HRTFs measured for various HATOs, Guldenschuh et al. [7] found that the most prominent torso reflections occur when ear, shoulder, and source are approximately aligned, and the source elevation is within 20° below the horizontal plane to 40° above. The authors further hypothesized that effects caused by the torso should be audible at least for critical source positions.

Despite the dominating role of head and pinnae effects on the HRTF, Genuit [8] assumed that the torso induces localization cues at frequencies below 3.5 kHz. This was supported with evidence by Algazi et al.[9]. Using 3kHz low-pass-filtered stimuli in localization experiments, the authors could show that torso cues indeed help in detecting the elevation of sound sources outside the median plane.

The studies discussed above support the hypothesis that accounting for correct HATO will be necessary for a perceptually transparent binaural synthesis. Yet, measuring HRTFs with high angular resolution and a large number of HATOs is time consuming making efficient methods for interpolation between different HATOs desirable. Various interpolation approaches were described for HRTFs obtained for different directions of sound incidence but constant HATO[10,11].

Hartung et al. [12] applied inverse distance weighting and spherical spline interpolation on HRIRs (time domain), and HRTFs log magnitude and phase spectra (frequency domain). Before applying interpolation in the time domain, HRIRs were time aligned on sample basis according to their maximum values (sampling rate 44.1 kHz). Inverse distance weighting is essentially a linear interpolation using a weighted average according to the great circle distance between the desired and actual source position, thus accounting for the spherical nature of HRTF data sets. When using spherical splines, interpolation is obtained by fitting polynomial functions to the data and evaluating them at the desired position given by azimuth and elevation. Smaller errors between interpolated and measured HRTFs were found for the frequency domain based methods with spherical spline interpolation tending to be superior to inverse distance weighting.

Using inverse distance weighting and minimum phase HRTFs, Minnaar et al. [13] investigated the minimum angular resolution needed for interpolating HRTFs without introducing audible artifacts. Physical evaluation revealed increasing interpolation errors for frequencies above 1 kHz. The largest errors were found at the contralateral ear, and at elevations below the horizontal plane, which is in good agreement with results of Hartung et al. [14]. Audibility of interpolation errors was assessed in a 3AFC listening test using a pink

[5] M. B. Gardner (**1973**). "Some monaural and binaural facets of median plane localization" *J. Acoust. Soc. Am.*

[6] V. R. Algazi, et al. (**2001**a). "Elevation localization and head-related transfer function analysis at low frequencies" *J. Acoust. Soc. Am.*

[7] M. Guldenschuh, et al. (**2008**). "HRTF modelling in due consideration variable torso reflections" in *Acoustics*.

[8] K. Genuit (**1984**). "Ein Modell zur Beschreibung von Außenohrübertragungseigenschaften" Doctoral Thesis.

[9] V. R. Algazi, et al. (**2001**a). "Elevation localization and head-related transfer function analysis at low frequencies" *J. Acoust. Soc. Am.*

[10] S. M. Robeson (**1997**). "Spherical methods for spatial interpolation: Review and evaluation" *Cartography and Geographic Information Systems*.

[11] R. Nicol (**2010**). *Binaural Technology*.

[12] K. Hartung, et al. (**1999**). "Comparison of different methods for the interpolation of head-related transfer functions" in *16th Int. AES Conference*.

[13] P. Minnaar, et al. (**2005**). "Directional resolution of head-related transfer functions required in binaural synthesis" *J. Audio Eng. Soc.*

[14] K. Hartung, et al. (**1999**). "Comparison of different methods for the interpolation of head-related transfer functions" in *16th Int. AES Conference*.

Source	1	2	3	4	5	6
Azim. φ_s [°]	0	315	0	45	90	315
Elev. ϑ_s [°]	90	30	0	0	0	-30
Distance [m²]	2.2	2.5	2.1	2.2	2.1	2.6

Table 3.1: Measured source positions.

noise stimulus, and covering directions of sound incidence from the horizontal, median and frontal plane. For most source positions, subjects failed to discriminate between measured HRTFs and HRTFs that were interpolated from a 4° grid. Occasionally differences remained detectable for lateral directions and below the horizontal plane.

Moreover, several studies transformed HRTF data sets into the spherical harmonic domain, where interpolation can be achieved by evaluating the spherical harmonic functions at the desired position given by azimuth and elevation[15,16,17].

In the present study, we physically and perceptually examined differences between dynamic auralizations of (a) HRTFs with constant and variable HATOs, as well as (b) measured and interpolated HRTFs. In the latter case, we specifically investigated the minimal resolution of HATOs required for interpolation artifacts to stay below the threshold of perception. We inferred that the torso effects should be most audible for head rotation to the left and right (termed *horizontal head rotations*), because in this case the largest changes of the ears' position relative to the torso occur. We hence limited our investigations accordingly.

[15] M. J. Evans, et al. (**1998**). "Analyzing head-related transfer function measurements using surface spherical harmonics" *J. Acoust. Soc. Am.*

[16] R. Duraiswami, et al. (**2004**). "Interpolation and range extrapolation of HRTFs" in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*.

[17] M. Pollow, et al. (**2012**). "Calculation of head-related transfer functions for arbitrary field points using spherical harmonics decomposition" *Acta Acust. united Ac.*

3.2 *Head-related transfer functions measurement*

Before being able to assess the effect of HATO, an appropriate HRTF data set was measured with the head and torso simulator FABIAN, which is equipped with a software-controlled neck joint, allowing for a precise control of the HATO in multiple degrees of freedom [18]. FABIAN's head and pinnae are casts of a human subject. The torso and the position of head and pinnae relative to the torso were designed according to anthropometric measures averaged across age and gender [19,20,21,22]. Accordingly, FABIAN's ear canal entrances are located 17.5 cm above and 1.5 cm in front of the acromion which is the highest point of the shoulder blade.

HRTFs were measured for six source positions given in Table 3.1. Thereby, azimuth angles $\varphi_s = \{0°, 180°, 90°, 270°\}$ denote sources in front and back, and to the left, and right of a listener's torso. Positive elevations ϑ_s denote sources above the horizontal plane. Accordingly, HATOs $\varphi_{HATO} = \{45°, 315°\}$ refer to a head rotation above the torso of 45° to the left, and right, respectively. Source position and HATO are independent, i.e. the source positions stays constant if the HATO changes and vice versa. Thus, torso-to-source

[18] A. Lindau, et al. (**2007**). "Binaural resynthesis for comparative studies of acoustical environments" in *122th AES Convention, Convention Paper 7032*.

[19] DIN 33402-2 (**2005**). *Ergonomics - Human body dimensions - Part 2: Values*.

[20] DIN IEC/TS 60318-7:2011 (**2005**). *Electroacoustics – Simulators of human head and ear – Part 7: Head and torso simulator for the measurement of hearing aids*.

[21] C. T. Morgan, et al. (**1963**). *Human Engineering Guide to Equipment Design*.

[22] K. Genuit (**1984**). "Ein Modell zur Beschreibung von Außenohrübertragungseigenschaften" Doctoral Thesis.

azimuth φ_{t2s} is given by $360 - \varphi_s$; the head-to-source azimuth φ_{h2s} by $(\varphi_{HATO} - \varphi_s)$ mod 360.

Source positions were chosen to be typical (e.g. on the horizontal plane) and particularly critical/non-critical with respect to a strong shoulder/torso effect and interpolation artifacts. Generally, source positions are critical for head orientations where ear, shoulder, and source are aligned (sources 2 to 5), this way giving rise to pronounced comb filters, or when the head and torso act as an obstacle for the sound field at the ears (sources 3 to 6), which results in strong shadowing at the contralateral ear, respectively. Source positions are less critical for sources well above the horizontal plane (source 1). Source distances between 2.1 m and 2.6 m were chosen to avoid proximity effects [23,24] and to ensure that reflections from the speakers could be removed by windowing.

The data set allowed the auralization of horizontal head rotations with constant and variable HATO within the physiological maximum range of motion $\varphi_{\text{HATO,max}} = \pm 82^\circ$ [25], and a resolution of $\Delta\varphi_{\text{HATO,ref}} = 0.5^\circ$. This spatial resolution is smaller than the worst-case localization blur of 0.75° reported by Blauert (p. 39)[26], and is termed *reference* in the following. Accordingly, 329 HRTFs for head rotations with constant, and 329 HRTFs for head rotations with variable HATO were measured for each source position. Moreover, additional HRTFs were measured to account for the different interpolation approaches. This will be described in more detail in Sec. 3.4.3 after introducing head and torso interpolation.

Measurements were conducted in the fully anechoic chamber of the TU Berlin ($V = 1850$ m^3, $f_c = 63$ Hz) using sine sweeps between 50 Hz and 21 kHz with an FFT order of 16 while achieving a peak-to-tail SNR of about 90 dB. FABIAN was mounted onto the turntable of a *VariSphear* microphone array (with the microphone removed) which gave high precision control of the torso-to-source orientation [27]. As sound sources we used *Genelec 8030a* active studio speakers with the tweeters aiming at FABIAN's interaural center (cross-over at 3 kHz, centers of tweeter and woofer 11 cm apart). Directivity measurements [28] showed that the major part of the torso laid within the speaker's main lobe for all source positions and frequencies ensuring that the effect of the torso is well represented in the measured HRTFs (cf. Fig 3.3). The time variability of the loudspeakers' frequency response could be reduced to ± 0.2 dB by means of an one-hour warming up procedure. The measurement setup is shown in Figure 3.2.

Subsequent to the HRTF measurements, FABIAN was removed and its *DPA 4060* miniature electret condenser microphones were detached for conducting reference measurements. The positions of the microphones were adjusted to be identical to FABIAN's interaural center. Finally, HRTFs were calculated by spectral division of the measured HRTF and the reference spectrum, simultaneously compensating for transfer functions of loudspeakers and microphones. Further processing of HRTFs included high-pass filtering for rejec-

[23] D. S. Brungart and W. M. Rabinowitz (**1999**). "Auditory localization of nearby sources. Head-related transfer functions" *J. Acoust. Soc. Am.*

[24] H. Wierstorf, et al. (**2011**). "A free database of head-related impulse resonse measurements in the horizontal plane with multiple distances" in *130th AES Convention, Enginieering Brief*.

[25] C. T. Morgan, et al. (**1963**). *Human Engineering Guide to Equipment Design*.

[26] J. Blauert (**1997**). *Spatial Hearing. The psychophysics of human sound localization*.

[27] B. Bernschütz (**2013**). "A spherical far field HRIR/HRTF compilation of the Neumann KU 100" in *AIA-DAGA 2013, International Conference on Acoustics*.

[28] J. G. Tylka, et al. (**2015**). "A database of loudspeaker polar radiation measurements" in *139th AES Convention, e-Brief 230*.

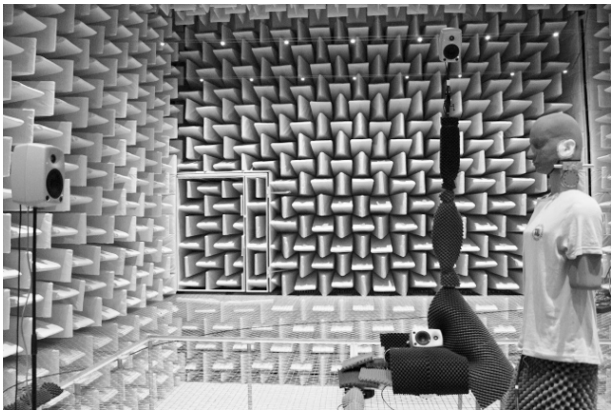


Figure 3.2: Photo of the HRTF measurement setup taken while adjusting the source position with the help of a laser mounted below FABIAN's left ear.

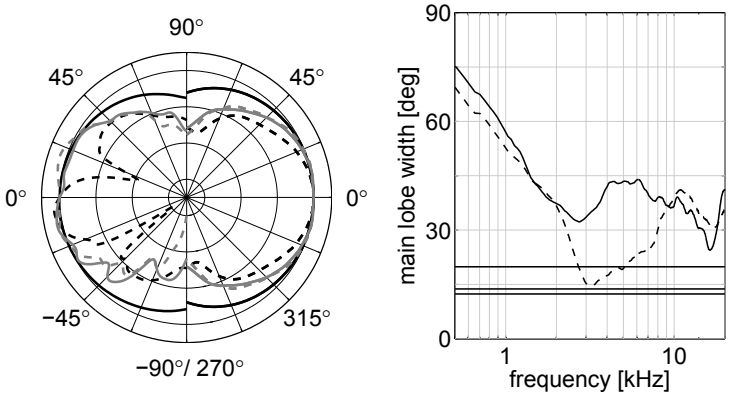


Figure 3.3: Left: Vertical and horizontal directivity (left and right semicircle) of a Genelec 8030a normalized at 0° (0.5 kHz, 3 kHz given by solid/-dashed black lines; 10 kHz, 20 kHz by solid/dashed gray lines. Grid spacing equals 10 dB). Right: Vertical (dashed) and horizontal (solid) main lobe width given by the angular distance between 0° and the -3 dB point. Horizontal lines mark FABIAN's shoulder-to-shoulder, ear-to-elbow, and ear-to-hip distance (46 cm; 51 cm; 76 cm) translated to an angular distance for a source at 2.1 m distance.

tion of low frequency noise, and shortening of the HRIRs to a length of 425 samples. Finally, HRIRs were saved as original phase, and minimum phase plus time of arrival (TOA) filters. Arrival times were estimated using onset detection on the ten times up-sampled HRIRs. Onsets were defined separately for the left and right channel by the first sample exceeding $\max\left(\left|HRIR_{l,r}\right|\right) - 6$ dB.

3.3 Effects of head-above-torso orientation

3.3.1 Physical evaluation

This section presents a physical evaluation of the torso's influence on HRTFs as a function of HATO and source position. Observed differences between head rotations with constant and variable HATO are discussed and a subset of source positions is selected for perceptual evaluation in a subsequent listening test.

Method: Differences in HRTFs were examined with respect to interaural time and level differences (ITD, ILD), as well as spectral fine structure. Therefore, ILDs were estimated as RMS level differences between left and right ear, whereas ITDs were calculated as differences in TOAs taken from the original phase HRIRs.

In order to obtain an impression of the spectral differences, the log-ratio of the magnitude responses between the HRTFs for constant

and variable head-above-torso conditions was calculated (in dB) as

$$\Delta\text{HRTF}(f) = 20\text{lg}\frac{|\text{HRTF}_{const}(f)|}{|\text{HRTF}_{var}(f)|}, \tag{3.1}$$

where f is the frequency in Hz. For convenience, the dependency of the HRTF on head orientation, source position, and left and right ear was omitted in (3.1)-(3.3).

For a better comparability across source positions, a single value measure was calculated based on Minnaar et al. [29], who described the error between a reference and an interpolated HRTF by averaging absolute magnitude differences at 94 logarithmically spaced frequencies, and adding results for left and right ear. This was found to be a good predictor for the listening test results in [30], where subjects had to detect differences between original and interpolated HRTFs. However, instead of calculating the error for discrete frequencies we used a Gammatone filter bank, as suggested by Schärer and Lindau [31]. The error level (in dB) in one filter band is given by

$$\Delta\text{HRTF}(f_c) = 20\text{lg}\frac{\int C(f, f_c)\,|\text{HRTF}_{const}(f)|\,df}{\int C(f, f_c)\,|\text{HRTF}_{var}(f)|\,df}, \tag{3.2}$$

where C is a Gammatone filter with center frequency f_c in Hz as implemented in the Auditory Toolbox [32]. The error level $\Delta\text{HRTF}(f_c)$ was calculated for $N = 39$ auditory filters between 70 Hz and 20 kHz. Then, the results for the left and right ear were added and averaged across f_c resulting in a single value error measure ΔG_μ (in dB) for each pair of HRTFs

$$\Delta G_\mu = \frac{1}{N}\sum_{f_c}\big(\,|\text{HRTF}_l(f_c)| + |\text{HRTF}_r(f_c)|\,\big). \tag{3.3}$$

Results: On average, ITD and ILD differences between head rotations with constant and variable HATO were found to be 2.6 μs, and 0.24 dB, and hence well below known difference thresholds (10 μs and 0.6 dB) from Blauert (1997, pp. 153)[33]. Maximum deviations of 11.4 μs and 0.95 dB exceeded assumed threshold levels only slightly.

HRTFs for head rotations with constant and variable HATO are depicted in Figure 3.4. In both cases, a comb-filter caused by the shoulder reflection is visible for frequencies above approx. 400 Hz. Above 3 kHz, it is partly masked by strong peak and notch patterns caused by pinnae resonances. However, when calculating the spectral difference according to (3.1) high frequency pinna cues cancel out due to identical head-to-source orientations. Expectedly, differences are nearly negligible for head orientations in the vicinity of 0°, as in this case head rotations with constant and variable HATO are very similar. For other head orientations comb-filter-like structures are visible from 0.4 to 20 kHz. In the cases of either constant or variable HATOs distances between ear and shoulder vary, resulting in 'detuned' comb filters whose differences can be seen in Figure 3.4. As a general trend, larger deviations occurred at the contralateral ear.

[29] P. Minnaar, et al. (**2005**). "Directional resolution of head-related transfer functions required in binaural synthesis" *J. Audio Eng. Soc.*

[30] P. Minnaar, et al. (**2005**). "Directional resolution of head-related transfer functions required in binaural synthesis" *J. Audio Eng. Soc.*

[31] Z. Schärer and A. Lindau (**2009**). "Evaluation of equalization methods for binaural signals" in *126th AES Convention, Convention Paper*.

[32] M. Slaney (**1998**). "Auditory toolbox. version 2" Technical Report #1998-010 .

[33] J. Blauert (**1997**). *Spatial Hearing. The psychophysics of human sound localization*.

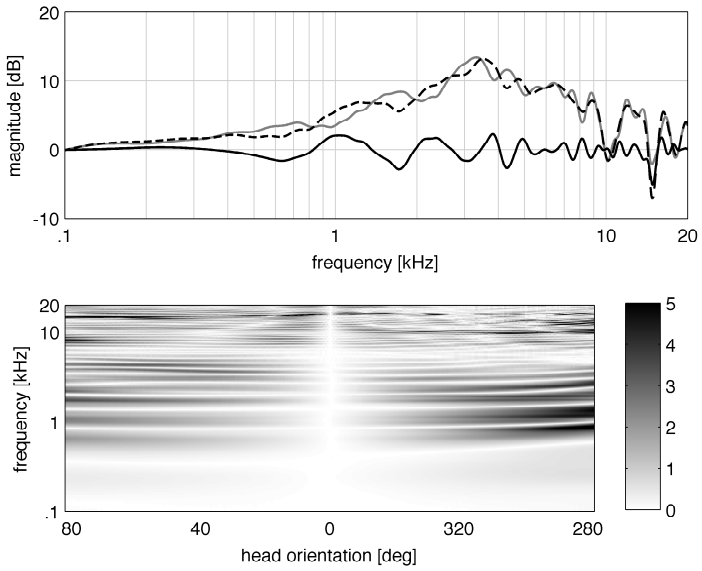


Figure 3.4: Right ear HRTFs of the source at (315°; 30°). Top: HRTF for a head orientation of 60° with constant (dashed) and variable (gray) HATO, and difference between them (black). Bottom: Difference between HRTFs with constant and variable HATO for all head orientations. Gray scale indicates magnitude in dB. Differences were calculated according to (3.1).

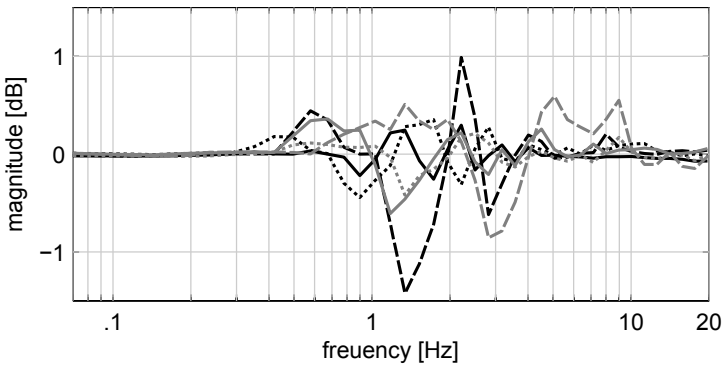


Figure 3.5: Differences between HRTFs with constant and variable HATO averaged across head orientations and left and right ear according to (3.2). Sources 1-3 are given by solid, dotted and dashed black lines; sources 4-6 by solid, dotted and dashed gray lines.

Below 700 Hz slight deviations can be seen which are probably due to shadowing effects of the torso. This finding is in good accordance with Algazi et al.[34], where strong shadowing was found for sound sources below $-40°$ elevation and the contralateral ear when using a KEMAR mannequin.

Spectral difference pattern according to (3.2) were comparable across sources and all exhibited comb-filter like structure (cf. Figure 3.5). It was thus assumed that the frequency independent measure according to (3.3) would give a fair impression of average differences for all source positions and head orientations (cf. Figure 3.6). Again, it can be seen that deviations are small in the vicinity of 0° whereas otherwise they reach a maximum of up to 2.4 dB. Moreover, a tendency for the error to increase with decreasing source elevation can be observed. The smallest error of ≤ 1 dB is found for the source at (0°; 90°). In this case, the shoulder reflection is weak for both constant and variable HATOs as most energy is reflected away from the ear. Intermediate differences of up to 1.4 dB occur for the sources on the horizontal plane and 30° elevation, most likely caused by strong shoulder reflections. The largest error of 2.4 dB is found for the source at $-30°$ elevation and for head-to-source orientations

[34] V. R. Algazi, et al. (**2002**). "Approximating the head-related transfer function using simple geometric models of the head and torso" *J. Acoust. Soc. Am.*

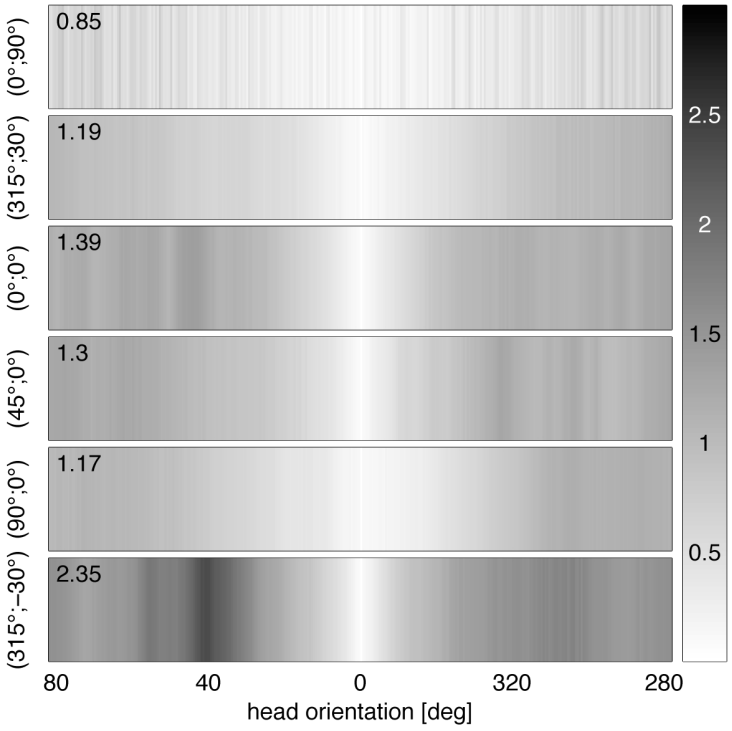


Figure 3.6: Differences between HRTFs with constant and variable HATO calculated according to (3.3). Values inside the plot indicate the maximum error per source. Gray scale indicates magnitude in dB.

larger than 45° azimuth, because the ear is partly shadowed by the torso in the case of constant HATO.

3.3.2 *Perceptual evaluation*

To test whether or not differences between head rotations with either constant or variable HATO are audible, an ABX listening test was conducted. The setup allowed for instantaneous and repeated comparison between HRTF sets using a dynamic binaural auralization accounting for horizontal head rotations of the listeners.

Method: Three women and eight men with a median age of 31 years took part in the listening test. All subjects had a musical background; ten subjects had participated in listening tests before; none reported known hearing impairments.

Following the ABX paradigm, three stimuli (A, B, and X) were presented to the subjects, whose task was to identify whether A or B equaled X. Conditions representing either head rotations with constant or variable HATO were randomly assigned to A, B, and X. Subjects were instructed and trained to listen to the stimuli in any order they felt to be helpful, to move/hold their heads to/at various positions during listening, to take their time at will before giving an answer, and to switch as fast or slow between stimuli as they wanted.

In order to limit the duration of the experiment, a subset of three sound sources was selected for perceptual evaluation. By drawing on the results of the physical evaluation, particularly critical and non-critical source positions at $(0°; 90°)$, $(90°; 0°)$, and $(315°; -30°)$ were selected. Two different audio stimuli were used: a frozen pink noise

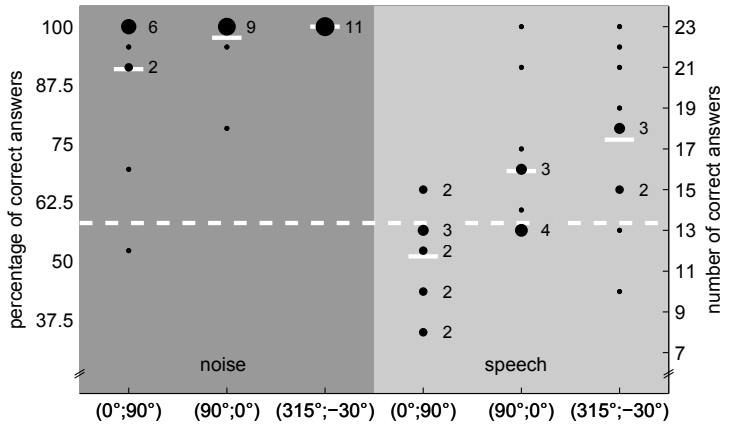


Figure 3.7: Listening test results for all subjects and conditions. Dots indicate percentage/number of correct answers; numbers indicate how many subjects had identical results (same number of correct answers). Group mean scores given by solid white lines above the dashed line are significantly above chance.

with a duration of 5 s (512 samples fade in/out) was chosen in order to reveal spectral differences, and an excerpt of German anechoic male speech with a duration of 5 s was used as a familiar and typical real-life sound. The experiment was split in two blocks whose sequence was balanced across subjects. Within a block, the source position was randomized while the audio content was held constant.

The combination of three sound sources and two audio contents lead to 2 x 3 = 6 conditions which were assessed individually by each subject. For each condition 23 ABX trials were conducted per subject, hence across subjects $23 \cdot 11 = 253$ trials were completed under each of the six conditions. Statistically, the test was designed to test a group averaged detection rate of 65% while guaranteeing cumulated type 1 and type 2 error levels to stay below 0.05 after accounting for repeated testing across conditions by Bonferroni correction [35]. Hence, for one tested condition detectability was significantly above chance when observing 147 or more correct answers.

For reproduction of binaural signals, a thoroughly evaluated dynamic auralization engine and dedicated extraaural headphones were used [36,37]. The test was conducted in a quiet listening room ($RT_{1\text{ kHz}} = 0.6\ s$; $V = 30\ \text{m}^3$; $L_{\text{eq,A}} = 33$ dB SPL), where subjects were seated on a revolving chair to comfortably reach and hold arbitrary head orientations. The listening test was administered using the whisPER environment [38], while displaying the user interface on a touchpad. Training prior to the listening test familiarized subjects with the interface and stimuli. Subjects were encouraged to take breaks at will to avoid fatigue, in turn needing maximally 1.5 hours for the test.

Results: Individual and group-averaged results are shown in Figure 3.7 for all tested conditions. Group-averaged results, as given by the white horizontal bars, indicate a clear distinguishability of head rotations with constant and variable HATO: Results were significantly above chance for all tested conditions, except for the non-critical source positions at $(0°; 90°)$ in conjunction with the speech stimulus. Moreover, significantly less correct answers were given for the speech stimulus ($\chi^2 = 44.66$, $p < 0.001$, $df = 1$). When asked for perceived differences between head rotations with constant and

[35] L. Leventhal (**1986**). "Type 1 and type 2 errors in the statistical analysis of listening tests" *J. Audio Eng. Soc.*

[36] A. Lindau and S. Weinzierl (**2012**). "Assessing the plausibility of virtual acoustic environments" *Acta Acust. united Ac.*

[37] F. Brinkmann, et al. (**2014**). "Assessing the authenticity of individual dynamic binaural synthesis" in *Proc. of the EAA Joint Symposium on Auralization and Ambisonics*.

[38] A. Lindau (**2014**). "Binaural resynthesis of acoustical environments. technology and perceptual evaluation" Ph.D. Thesis.

variable HATO, the subjects mentioned coloration (11x) and/or localization (3x) in the case of the noise content, and coloration (6x), localization (5x), and/or source width (1x) for the speech sample.

So far, we discussed differences between head rotations with constant and variable HATO. For a number of different conditions we could show that the acoustic deviations between these two situations are audible. We thus conclude that variable HATOs have to be considered when aiming at a perceptually transparent binaural synthesis. In the remainder of this paper, we will discuss interpolation approaches suitable for an efficient representation of HATO in acoustic simulations.

3.4 Interpolation of head-above-torso orientation

In this section we first introduce and discuss different approaches to spatial interpolation of HRTFs. Second, we show a physical evaluation of in total 17 individual interpolation algorithms. Finally, we present the perceptual evaluation of a selected subset of these algorithms, and extend the results towards all approaches based on a perceptually motivated error measure.

3.4.1 Inverse distance weighting and spline interpolation

Interpolation algorithms for spherical data such as HRTFs may be distinguished with respect to wether they operate on neighboring data points only (nearest neighbor, inverse distance weighting, polynomials, splines), or whether they require a full-spherical data set (spherical splines, spherical harmonics). Nevertheless, in principle both families of approaches could be used for the interpolation of HATO. In the latter case however, spherical spline or spherical harmonic coefficients had to be interpolated instead of directly interpolating HRTFs. Consequently, when aiming at finding the difference threshold, full spherical HRTF data sets for HATOs between $\pm 82^\circ$ in the smallest resolution $\Delta\varphi_{\mathrm{HATO,meas}} = 1^\circ$ were needed for calculation of the corresponding coefficients and successive interpolation onto the reference $\Delta\varphi_{\mathrm{HATO,ref}}$. As this would require an unfeasibly large amount of measured data, the current study was restricted to spline – instead of spherical spline – interpolation, and inverse distance weighting. Moreover, interpolation was applied in the time and frequency domain as well as for original and minimum phase HRTFs.

Depending on the head orientation, HRIRs contain different arrival time delays. As a consequence, neighboring HRIRs are temporally misaligned and a direct time domain interpolation would result in double/blurred peak HRIRs. Two alignment strategies were applied to overcome this problem. On the one hand, arrival times were estimated using onset detection as described in Sec. 3.2. On the other hand, we estimated the amount of misalignment from the

cross-correlation function between two ten times up-sampled HRIRs $\left(\arg\max_\tau \varrho_{xy}(\tau)\right)$. In both cases fractional delays were applied for time alignment [39]. Additionally, TOAs were interpolated based on the the extracted values for both alignment procedures. In the frequency domain, magnitude and unwrapped phase spectra of the original phase HRTFs were interpolated separately, thus again inherently interpolating the TOA and ITD. For minimum phase HRIRs only the magnitude spectrum was interpolated and the result was made minimum phase again using the Hilbert transformation [40]. In this case, the TOA had to be interpolated separately for both time and frequency domain interpolation.

[39] T. I. Laakso, et al. (**1996**). "Splitting the unit delay" *IEEE Signal Processing Magazine*.

[40] A. V. Oppenheim, et al. (**1999**). *Discrete-time signal processing*.

In addition to spline interpolation and inverse distance weighting, the nearest neighbor method was applied. In this case, the HRIR either with the HATO closest to the target orientation (similar to head interpolation) or with the closest torso-to-source azimuth (similar to torso interpolation) was used. This method was included as a possible approach for applications with limited computational resources as, e.g. in mobile applications. Because the nearest neighbor method yields identical results in the frequency and time domain, as well as for original and minimum phase HRIRs, only one variation had to be tested. In total, 17 interpolation algorithms were investigated as listed in Table 3.2, and described in more detail in the following.

With inverse distance weighting, HRTFs for intermediate HATOs φ'_{HATO}, source azimuth φ'_{s}, and elevation ϑ'_{s} are obtained as a weighted average of neighboring positions

$$x\left(\varphi'_{\mathrm{HATO}}, \varphi'_{\mathrm{s}}, \vartheta'_{\mathrm{s}}\right) = \frac{\sum_{i=1}^{2} x\left(\varphi_{\mathrm{HATO,i}}, \varphi_{\mathrm{s,i}}, \vartheta_{\mathrm{s,i}}\right) d^{-1}_{\varphi,\varphi'}}{\sum_{i=1}^{2} d^{-1}_{\varphi,\varphi'}} \tag{3.4}$$

whereby x denotes a sample of the HRIR in the case of time domain interpolation, and a bin of the HRTFs magnitude or phase response in the case of frequency domain interpolation. For head rotations restricted to the horizontal plane, the great circle distance $d_{\varphi,\varphi'}$ reduces to

$$d_{\varphi,\varphi'} = \arccos\left(\cos\left(\varphi_{\mathrm{HATO,i}} - \varphi'_{\mathrm{HATO}}\right)\right). \tag{3.5}$$

The neighboring HATOs are given by

$$\varphi_{\mathrm{HATO,i}} = \left[\left(\left\lfloor \frac{\varphi'_{\mathrm{HATO}}}{\Delta\varphi_{\mathrm{HATO,meas}}} \right\rfloor + \mathrm{i}\right)\Delta\varphi_{\mathrm{HATO,meas}}\right] \bmod 360 \tag{3.6}$$

where $\mathrm{i} \in \{0, 1\}$, $\lfloor\cdot\rfloor$ denotes rounding to the next lower integer, and *mod* is the modulus operator. In contrast, cubic spline interpolation fits a piecewise polynomial through all $x\left(\varphi_{\mathrm{HATO,i}}\right)$ with a continuous first and second derivate on the entire interval [41], whereby

[41] W. Gautschi (**2012**). *Numerical analysis*.

$$\varphi_{\mathrm{HATO,i}} = \left(\mathrm{i}\,\Delta\varphi_{\mathrm{HATO,meas}}\right) \bmod 360, \tag{3.7}$$

with $-N \leq \mathrm{i} \leq N$, $i \in \mathbb{Z}$, and $N = \lceil 82^\circ / \Delta\varphi_{\mathrm{HATO,meas}} \rceil$.

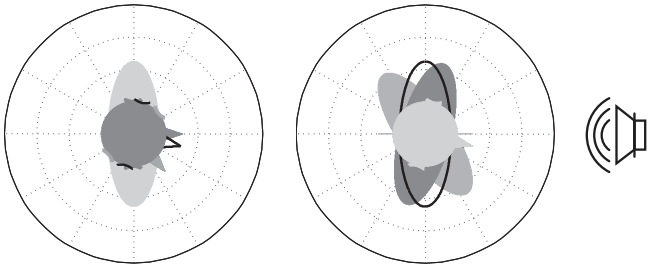


Figure 3.8: Illustration of head (left) and torso interpolation (right) for inverse distance weighting ($\Delta\varphi_{\mathrm{HATO,meas}} = 50°$, $\varphi'_{\mathrm{HATO}} = 350°$, $\varphi'_{\mathrm{s}} = 0°$). Positions of measured HRTFs are shown with solid heads and torsi, interpolated HRTFs are indicated by black lines.

3.4.2 *Head and torso interpolation*

Two different approaches can be considered when interpolating HATO in HRTFs. With head interpolation, intermediate data points are calculated from HRTFs with identical torso-to-source but differing head-to-source orientations (Figure 3.8, left). Thus, HRTFs used for interpolation will deviate primarily in the high frequency range, which is dominated by direction-dependent (anti) resonance effects of the pinnae cavities. Hence, this approach is comparable to interpolating HRTFs of different sound source positions and thresholds are expected to be in the order given by Minnaar[42].

In the case of torso interpolation, HRTFs with identical head-to-source but differing torso-to-source orientations are used for the estimation of intermediate points (Figure 3.8, right). This approach appears promising because the spectral effect of the torso in HRTFs is less prominent for most directions of sound incidence and the dominating high frequency structure will remain preserved. However, it requires additional HRTFs with source azimuths $\varphi_{\mathrm{s,i}}$ for interpolating the desired source azimuth φ'_{s}

$$\varphi_{\mathrm{s,i}} = (\varphi_{\mathrm{HATO,i}} - \varphi'_{\mathrm{HATO}} + \varphi'_{\mathrm{s}}) \bmod 360, \tag{3.8}$$

where i and $\varphi_{\mathrm{HATO,i}}$ remain as specified for Eq. (3.6-3.7). As depicted in Figure 3.8, Eq. (3.8) ensures that the head-to-source azimuth remains constant while the torso is rotated with respect to the source resulting in a change of φ_s. When applying inverse distance weighting to torso interpolation, two additional HRTFs with differing source azimuths are needed for each interpolation, whereas spline interpolation would require a multitude of additional HRTFs. Although this is not a drawback in practice as HRTF data sets usually cover source positions in a high spatial resolution, spline interpolation was excluded from this study in the case of torso interpolation due to the increased measurement effort. Nevertheless, we hypothesized that interpolation artifacts are smaller for torso interpolation compared to head interpolation.

[42] P. Minnaar, et al. (**2005**). "Directional resolution of head-related transfer functions required in binaural synthesis" *J. Audio Eng. Soc.*

3.4.3 *Additional head-related transfer function measurements*

Head and torso interpolation were investigated for 23 different resolutions of measured HATOs $\Delta\varphi_{\mathrm{HATO,meas}} = \{1,2,\ldots,10,12,\ldots,30,35,\ldots,45°\}$ in the range of $\pm 82°$ given by $\varphi_{\mathrm{HATO,max}}$. Hence, additional HRTFs had to be measured: First, they were needed in cases where

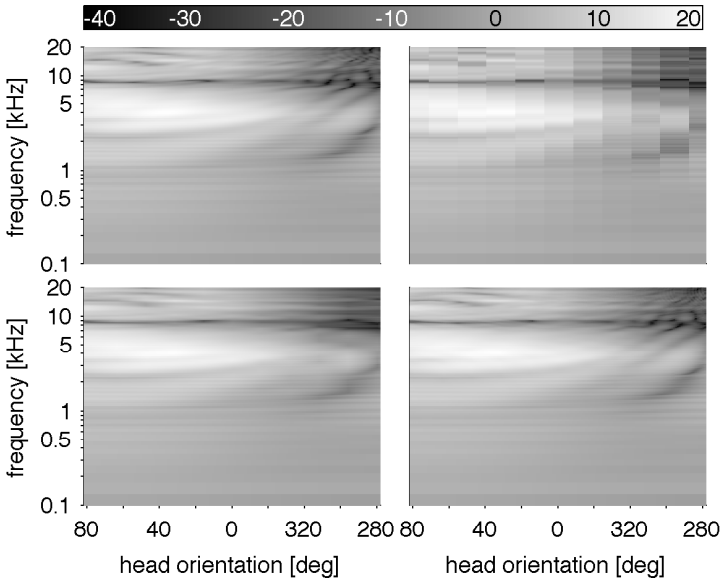


Figure 3.9: Magnitude spectra of reference (top left) and interpolated HRTFs (linear interpolation, time domain, $\Delta\varphi_{\mathrm{HATO,meas}} = 16°$, source 3, right ear). Nearest neighbor, via head (top right); head interpolation (bottom left); torso interpolation (bottom right). Gray scale indicates magnitude in dB.

$\Delta\varphi_{\mathrm{HATO,meas}}$ was not an integer divisor of $\varphi_{\mathrm{HATO,max}}$. For example, HATOs of 60° and 90° were needed to interpolate to 82°, in the case of $\Delta\varphi_{\mathrm{HATO,meas}} = 30°$. Second, additional HRTFs were needed for testing torso interpolation: Because the torso is rotated during interpolation, two additional source positions had to be measured for each intermediate HATO, i.e. if φ'_{HATO} is not an integer divisor of $\Delta\varphi_{\mathrm{HATO,meas}}$ (cf. Eq. (3.8), and Fig 3.8). This lead to $\sum_{\forall k} 2 \cdot (329 - (2 \cdot \lfloor \varphi_{\mathrm{HATO,max}} / \Delta\varphi_{\mathrm{HATO,meas,k}} \rfloor + 1))$ additional HRTFs. Calculating the corresponding HATOs and source positions using (3.6)-(3.8) and removing duplicates resulted in 6679 additional HRTFs that were measured for each sound source listed in Table 3.1.

3.4.4 *Physical evaluation*

A physical evaluation of all 17 algorithms was carried out, calculating differences between the reference and interpolated HRTFs according to (3.1). For this purpose, HRTFs were interpolated in the range of $-82° \leq \varphi'_{\mathrm{HATO}} \leq 82°$ to $\Delta\varphi_{\mathrm{HATO,ref}} = 0.5°$ for each measured resolution $\Delta\varphi_{\mathrm{HATO,meas}}$ and algorithm (cf. Figure 3.9). Whereas in the reference, smallest changes in the high frequency fine structure are smoothly reproduced, discontinuities are clearly seen for the nearest neighbor algorithm, due to the hard switching between impulse responses for discrete HATOs. When comparing head interpolation to the reference, impairments become visible above approximately 2 kHz. In contrast, with torso interpolation, the spectral fine structure is mostly preserved.

Differences between reference and interpolated HRTFs according to (3.2) are shown in Figure 3.10. They confirm our hypothesis that the errors for torso interpolation are smaller then for head interpolation. In general, and in accordance to Minnaar et al.[43], errors increase with frequency which is most likely related to high frequency pinnae cues in the HRTF that underlie a fast spatial fluctuation. Due

[43] P. Minnaar, et al. (**2005**). "Directional resolution of head-related transfer functions required in binaural synthesis" *J. Audio Eng. Soc.*

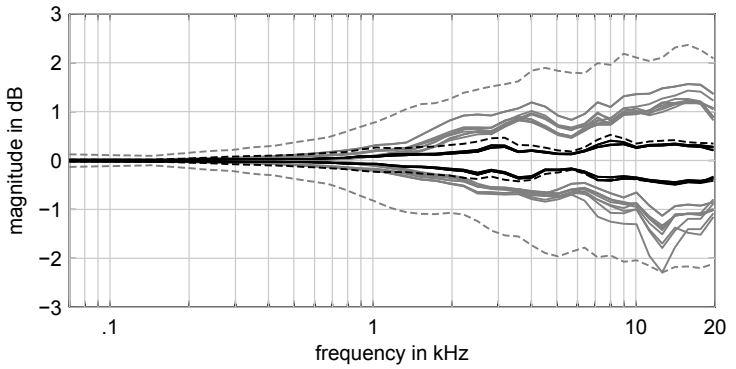


Figure 3.10: 5-95% percentile range of the error between reference and interpolated HRTFs according to (3.2) for all interpolation algorithms (averaged across sources, head orientations, and $\Delta\varphi_{\mathrm{HATO,meas}}$). Gray lines show head, black lines torso interpolation; dashed lines refer to the nearest neighbor approach.

#	Approach	Interp.	Domain	Phase	Alignm.	(0°;90°)	(315°;30°)	(0°;0°)	(45°;0°)	(90°;0°)	(315°;-30°)	*mean*
1	head interpolation	nearest	time	org.	–	0.14	0.63	0.77	0.83	0.83	0.67	***0.64***
2		linear	time	org.	cross cor.	0.11	0.21	0.29	0.33	0.35	0.22	***0.25***
3					ons.	0.11	0.22	0.30	0.35	0.37	0.24	***0.26***
4				min.	–	0.11	0.21	0.28	0.31	0.34	0.22	***0.25***
5			freq.	org.	–	0.11	0.18	0.24	0.27	0.30	0.19	***0.22***
6				min .	–	0.11	0.19	0.24	0.27	0.30	0.19	***0.22***
7		spline	time	org.	cross cor.	0.12	0.14	0.20	0.23	0.25	0.15	***0.18***
8					ons.	0.12	0.14	0.21	0.24	0.25	0.16	***0.19***
9				min.	–	0.12	0.15	0.20	0.23	0.25	0.16	***0.19***
10			freq.	org.	–	0.11	0.12	0.17	0.21	0.22	0.13	***0.16***
11				min.	–	0.12	0.12	0.17	0.20	0.22	0.13	***0.16***
12	torso interpolation	nearest	time	org.	–	0.14	0.11	0.16	0.18	0.10	0.26	***0.16***
13		linear	time	org.	cross cor.	0.11	0.07	0.09	0.11	0.07	0.16	***0.10***
14					ons.	0.11	0.07	0.10	0.11	0.07	0.17	***0.10***
15				min.	–	0.12	0.07	0.10	0.11	0.07	0.17	***0.11***
16			freq.	org.	–	0.11	0.07	0.08	0.10	0.06	0.14	***0.09***
17				min.	–	0.11	0.07	0.08	0.10	0.06	0.14	***0.09***

Table 3.2: Median error between reference and interpolated HRTFs according to (3.3) for all interpolation algorithms and source positions (averaged across head orientations and $\Delta\varphi_{\mathrm{HATO,meas}}$). Means across sources are given for ease of interpretation. Errors of 0 dB that occur when the head orientation is a multiple of $\Delta\varphi_{\mathrm{HATO,meas}}$ were excluded from analysis.

to the similarity of the error pattern, it was again assumed that (3.3) still reflects differences between interpolation algorithms and source positions. For an overview of the average performance of algorithms and source positions, median errors averaged across HATO and $\Delta\varphi_{\mathrm{HATO,meas}}$ are given in Table 3.2.

Differences between approaches follow the line of argumentation given above. When looking at results for head interpolation, a slight superiority of spline compared over linear interpolation can be seen (0.18 vs. 0.24 dB on average). If excluding the nearest neighbor approach, results for time and frequency domain interpolation as well as for time alignment by cross correlation and onset detection are comparable. In tendency however, smaller errors occur in the frequency domain (0.16 vs. 0.18 dB) and when using cross correlation (0.18 vs. 0.19 dB). Moreover, average performance for original and minimum phase processing (0.17 dB), as well as for the best head interpolation compared to torso interpolation when using the nearest neighbor approach (0.16 dB) were identical.

Results for the source positions depend on the interpolation approach. For head interpolation, errors are largest for sources on the

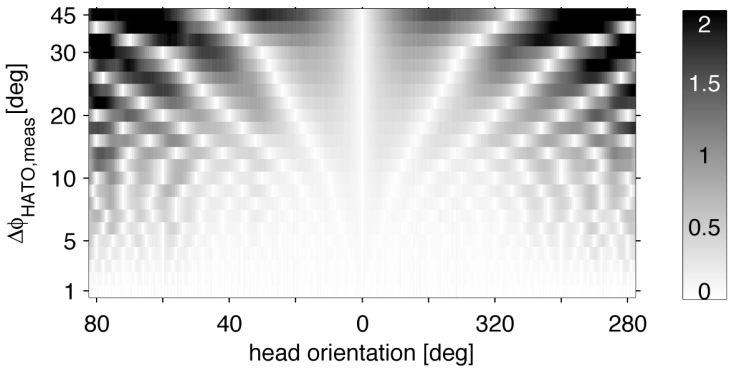


Figure 3.11: Example of errors between reference and interpolated HRTFs according to (3.3): Original phase, time domain, head interpolation with cross correlation for HRIR alignment.

horizontal plane, slightly smaller for sources at $\pm 30^\circ$ elevation, and smallest for the source at 90° elevation. Interestingly, this rank order is exactly reflected in the median ILD per source and across head orientations (not shown here). This is in agreement with Hartung and Minnaar[44,45] who reported interpolation errors to increase with increasing source-to-head azimuth (i.e. with increasing ILD) due to a lower SNR at the contralateral ear. For torso interpolation, in general, differences between sources are smaller. Largest errors occurred for source 6 and smallest for source 5. If averaged across all algorithms, errors for source 1 are smallest and almost identical, indicating its non-critical nature towards interpolation artifacts. Moreover, errors for source 6 are comparable for torso interpolation (0.16 dB @ linear interp.) and head interpolation (0.15 dB @ spline interp.). However, a more detailed analysis revealed that this only holds for $\Delta\varphi_{\mathrm{HATO,meas}} \lesssim 10^\circ$, otherwise, torso interpolation exhibits smaller errors (0.31 dB vs. 0.37 dB).

As an example for one interpolation approach, results according to (3.3) are shown in Figure 3.11 for all head orientations and $\Delta\varphi_{\mathrm{HATO,meas}}$. As expected, errors were zero at multiples of $\Delta\varphi_{\mathrm{HATO,meas}}$, largest in between, and increased with increasing measurement grid width.

3.4.5 *Perceptual evaluation*

Method: Difference thresholds – defined as the inflection point of the sigmoid psychometric function – between reference and interpolated HRTFs were determined using a parametric, adaptive three alternative forced choice test utilizing the ZEST adaptive procedure. ZEST provides a fast and unbiased threshold estimation, which is robust against uncertainties with respect to its proper parameterization [46,47]. The test was parameterized with a logistic psychometric function (slope parameter $\beta = 1$), a Gaussian a priori probability density function (mean set according to informal listening tests, standard deviation set to $\sigma = 25$), and a lapsing rate of 3%. Again, the whisPER listening test environment was used for conducting the experiment.

Following a 3AFC paradigm, the subjects' task was to detect the interpolated HRTFs by finding the oddball in three presented stimuli. Subjects were carefully instructed and trained to listen to the

[44] K. Hartung, et al. (**1999**). "Comparison of different methods for the interpolation of head-related transfer functions" in *16th Int. AES Conference*.

[45] P. Minnaar, et al. (**2005**). "Directional resolution of head-related transfer functions required in binaural synthesis" *J. Audio Eng. Soc.*

[46] B. Treutwein (**1995**). "Adaptive psychophysical procedures" *Vision Research*.

[47] S. Otto and S. Weinzierl (**2009**). "Comparative simulations of adaptive psychometric procedures" in *NAG/DAGA 2009, International Conference on Acoustics*.

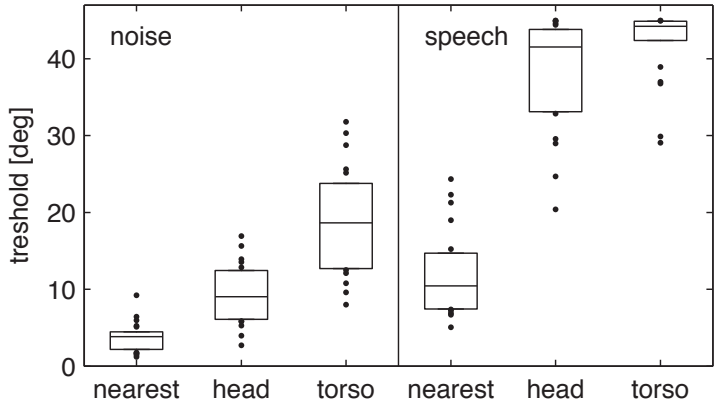


Figure 3.12: Distribution of difference thresholds (in degree) for all conditions and subjects. Boxes show median and interquartile range (IQR). Values outside the IQR are marked by dots.

stimuli in any order they felt to be helpful, to move/hold their heads to/at various positions during listening, to take their time at will before giving an answer, and to switch as fast or slow between stimuli as they wanted. This was important, because the spatial regions of largest interpolation errors strongly depend on the interpolation interval, which changed continuously during the adaptive test procedure. Dynamic auralization for HRTFs with HATOs between $\pm 82^\circ$ was realized as described in Sec. 3.3.2.

Two types of audio stimuli (continuous pink noise; anechoic male speech) and three algorithms ([A] nearest neighbor, via head; [B] head interpolation: time domain, spline interpolation; [C] torso interpolation: frequency domain, linear interpolation) were tested in a two-way factorial, fully repeated measures design (2 x 3 = 6 conditions per subject). The experiment was conducted in an acoustically dry recording studio environment ($RT_{1\ \mathrm{khz}} = 0.5$ s; $V = 145\ \mathrm{m}^3$; $L_{\mathrm{eq,A}} =$ 23 dB SPL).

In order to limit the listening test duration, only the source position most critical towards torso interpolation ($315^\circ; -30^\circ$) was tested using minimum-phase HRTFs. To avoid listening fatigue, the test was split in two blocks, each starting with a training followed by three threshold estimates (20 trials each) and an intermediate break of 30 minutes or more. The presentation order of audio stimuli and algorithms was balanced across subjects, while the stimulus was held constant within blocks.

Results: Thresholds for 25 subjects (6 women, 19 men, median age 27, 24 subjects had musical background, 22 participated in listening tests before) are shown in Figure 3.12. Two subjects were discarded from statistical analysis because they were short on time and hurried to finish the test. In turn, both subjects rated noticeably faster than others while showing considerably worse results.

Statistical analysis by means of ANOVA requires normally distributed samples. Because this criterion was violated under some conditions (Lilliefors test), non-parametrical tests were used for analyzing the results. Friedman's test showed highly significant differences between conditions ($\chi^2 = 112.4$, $p < .001$) Hence, as hypothesized, detectability thresholds increase from the nearest neighbor

approach to torso interpolation when pooled across stimuli. Additionally, thresholds were higher for the speech as compared to the pink noise stimulus when pooled across algorithms. Post-hoc pairwise comparisons proved all observed differences to be highly significant (Wilcoxon signed rank tests, $p < 0.001$ after accounting for multiple testing by means of Bonferroni correction).

From inspection of the subjects' answers, we assumed that some did not hear differences between reference and interpolated HRTFs regardless of the interpolation interval when being presented with the speech stimulus. To support this assumption, Bernoulli tests [48] were carried out based on answers obtained for the largest measurement grid of $\Delta\varphi_{\mathrm{HATO,meas}} = 45°$. They revealed that one (head interpolation), and eight (torso interpolation) subjects failed to significantly discriminate between reference and interpolated HRTFs (type 1 and 2 error 0.025, Bonferroni corrected for multiple testing; testable effect $p = 0.9$). Keeping in mind that the presentation order was balanced and that all subjects detected differences for the noise stimulus, this was believed to be solely related to the speech signal. Its non-stationary and band limited nature made it harder to detect differences, which apparently were below individual thresholds of these subjects for all grid widths. Consequently, we assumed the measured threshold to be underestimated in this case because of this described ceiling effect.

[48] L. Leventhal (**1986**). "Type 1 and type 2 errors in the statistical analysis of listening tests" *J. Audio Eng. Soc.*

In order to recommend the required measurement grid size that is needed to achieve or fall below a given group-averaged detectability, cumulated probability density functions were estimated from subjects' thresholds using a non-parametric modeling algorithm [49]. Grid width $\Delta\varphi_{\mathrm{HATO}}$ for selected percentiles of average detectability are listed in Table 3.3, and will be referred to as *threshold percentiles* in the following. Thereby, for example, the 5% threshold percentile denotes the grid width that is below the threshold of perception for 95% of the population underlying the subjects that participated in the listening test. For the noise stimulus, threshold percentiles increase by a factor of approximately two across algorithms, suggesting that differences between them are perceptually relevant. For the speech stimulus, this factor is even larger from nearest neighbor to head interpolation, but due to the ceiling effect small between head and torso interpolation.

[49] K. Żychaluk and D. H. Foster (**2009**). "Model-free estimation of the psychometric function" *Attention, Perception, & Psychophysics*.

When asked for perceived differences, subjects mentioned coloration (23x), and localization (13x) in the case of the noise stimulus, and localization (20x), and coloration (16x) for the speech sample.

3.4.6 Threshold prediction

To extend the results obtained in the perceptual evaluation towards interpolation algorithms that were not included in the listening test, thresholds for all algorithms and source positions were predicted based on an investigation of the interpolation error in dependency of the grid width $\Delta\varphi_{\mathrm{HATO,meas}}$. According to (3.2), we obtained one

		Noise			Speech		
		Near.	**Head**	**Torso**	**Near.**	**Head**	**Torso**
50%	$\Delta\varphi_{\text{HATO}}$	4.4	10.5	20.8	12.7	41.1	43.6
	ΔG_{95}	0.89	1.09	0.82	2.15	3.64	1.28
25%	$\Delta\varphi_{\text{HATO}}$	3.3	7.9	15.5	9.5	35.6	41.4
	ΔG_{95}	0.69	0.71	0.68	1.74	2.81	1.24
5%	$\Delta\varphi_{\text{HATO}}$	1.5	4.6	9.6	5.7	23.7	31.7
	ΔG_{95}	0.41	0.33	0.44	1.12	2.44	1.12

Table 3.3: Threshold percentiles $\Delta\varphi_{\text{HATO}}$, and corresponding error vales ΔG_{95} (cf. Sec. 3.4.6) for all tested conditions.

error measure per HATO and auditory filter, resulting in

$$\text{329 HATOs x 39 audit. filter } = 12,831\Delta\text{HRTF}(f_c)$$

values for each grid width and source position. By assuming that (a) differences between reference and interpolated HRTFs are audible if any $\Delta\text{HRTF}(f_c)$ exceeds a certain threshold, and (b) that due to the dynamic auralization the highest $\Delta\text{HRTF}(f_c)$ might not always be discovered, we expected the arithmetic mean across the largest five percent of the 12,831 values to be a perceptually suitable and robust error measure. This measure was termed ΔG_{95} and is depicted in Figure 3.13. Expectedly, head interpolation exhibits larger errors than torso interpolation, and the nearest neighbor approach represents the upper error bound except for grid widths larger than 40°, where occasionally errors are largest for spline interpolation. In general, the error increases with increasing grid width, but especially for head interpolation local maxima and minima emerge. This indicates that the quality of interpolation is not only a function of grid width.

To establish a link to the results for source $(315°; -30°)$ obtained from perceptual evaluation, ΔG_{95} was calculated at the 5%, 25%, and 50% threshold percentiles given in Table 3.3. If $\Delta\varphi_{\text{HATO}}$ was not included in the measured HATO resolution $\Delta\varphi_{\text{HATO,meas}}$, ΔG_{95} was calculated using a weighted average of the two neighboring values. For the noise stimulus the ΔG_{95} values as expected (a) are approximately equal within a given threshold percentile, (b) do not overlap across threshold percentiles, and (c) decrease with decreasing threshold percentile. In this cases this indicates their perceptual relevance, and their suitability to be used for predicting threshold percentiles for sources and interpolation algorithms that were not included in the perceptual evaluation. This is, however, not the case for the speech stimulus which might either be caused by the ceiling effect that biased the threshold percentiles in Table 3.3, or it might suggest that the pure spectral error measure ΔG_{95} loses its validity in this case.

Finally, for predicting the threshold percentiles for all interpolation approaches, only the ΔG_{95} values obtained for the noise stimulus were used. For robustness, ΔG_{95} was averaged across the tested in-

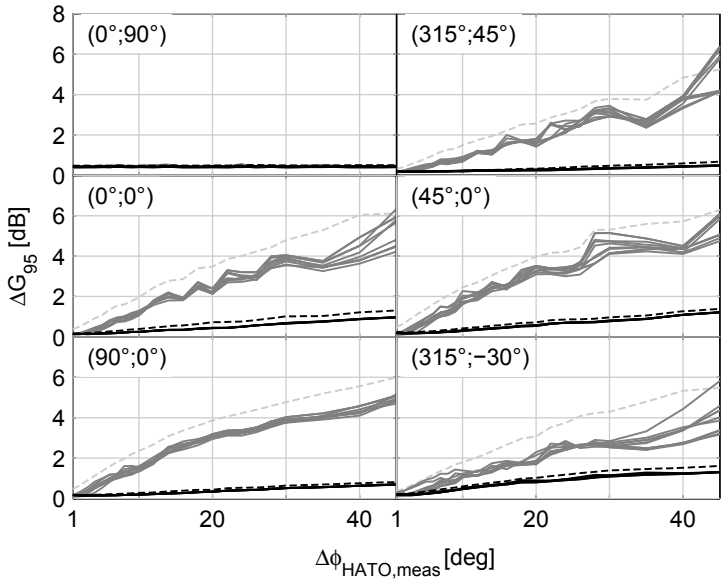


Figure 3.13: ΔG_{95} for all source positions and grid widths $\Delta\varphi_{\mathrm{HATO,meas}}$. Gray lines show head, black lines torso interpolation; dashed lines refer to the nearest neighbor approach.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
50%	3.1	5.0	4.6	5.6	5.7	5.7	6.5	6.2	6.5	6.2	6.3	17.2	23.6	20.9	23.6	25.1	25.2
25%	2.1	3.9	3.5	4.3	4.5	4.5	5.6	4.2	4.9	5.0	5.0	11.9	14.9	14.4	14.7	15.9	16.0
5%	1.0	2.8	2.1	2.8	3.0	3.0	4.4	2.0	3.5	3.6	3.6	5.8	8.6	7.6	8.3	8.7	8.7

Table 3.4: Threshold estimates in degree for all 17 interpolation algorithms. For ease of display, only the smallest (most critical) estimate across the six source positions is shown. The interpolation algorithms are numbered according to Table 3.2 (1: head interp., near. neighb.; 2-11: head interp., lin./spline; 12 torso interp., near. neighb.; 13-17: torso interp., lin./spline).

terpolation algorithms, which lead to the following values that were used for prediction: $\Delta G_{95,\mathrm{pred.},50\%} = 0.93$ dB, $\Delta G_{95,\mathrm{pred.},25\%} = 0.69$ dB, and $\Delta G_{95,\mathrm{pred.},5\%} = 0.39$ dB. Subsequently, thresholds percentiles for all interpolation algorithms and sources were predicted by finding the first ΔG_{95} value exceeding the corresponding $\Delta G_{95,\mathrm{pred.}}$. To make this prediction more exact, the curves in Figure 3.13 were interpolated to a resolution of 0.01° beforehand.

Noteworthy, the average difference between the nine threshold percentiles estimated from the perceptual evaluation and the predicted threshold percentiles was only 0.9°. However, a deviation of 5° between thresholds occurred for the torso interpolation (algorithm #17, source 6, 50% percentile). This was caused by the slow increase of ΔG_{95} for $18° \leq \Delta\varphi_{\mathrm{HATO}} \leq 25°$ (cf. Figure 3.13) and was thus considered to be perceptually non-critical. The smallest predicted threshold percentile for each interpolation algorithm across sources is listed in Table 3.4.

3.5 *Discussion*

Our evaluation of the effect of HATO in HRTFs supported findings of earlier studies regarding the comb-filter like nature of the shoulder reflection, which was found to be most prominent if sound source, shoulder, and ear are aligned [50,51,52]. Because observed deviations in ITDs and ILDs were below the threshold of audibility for the vast majority of HATOs and source positions, we suppose perceived differ-

[50] M. Guldenschuh, et al. (**2008**). "HRTF modelling in due consideration variable torso reflections" in *Acoustics*.

[51] V. R. Algazi, et al. (**2002**). "Approximating the head-related transfer function using simple geometric models of the head and torso" *J. Acoust. Soc. Am.*

[52] V. R. Algazi, et al. (**2001a**). "Elevation localization and head-related transfer function analysis at low frequencies" *J. Acoust. Soc. Am.*

ences to be mostly due to spectral deviations. Perceived differences in localization might also be due to spectral cues, related to mismatched comb-filters in HRTFs exciting different directional bands (cf. Blauert, pp. 93[53]) and thus evoking differences in perceived elevation. This assumption would be in accordance with Algazi et al. [54], who found torso and shoulder related cues to be involved in the perception of elevation for sources outside the median plane.

Best interpolation results were – as presumed a priori and predicted by physical evaluation – achieved for torso interpolation using HRTFs with identical head-to-source but varying torso-to-source orientation. Compared to head interpolation, this provided a better preservation of high frequency pinnae cues when interpolating between HRTFs with identical head-to-source orientation. Remarkably, torso interpolation in conjunction with the nearest neighbor approach outperformed most head interpolation algorithms, thus suggesting that the effect of the torso on the HRTF is small compared to that of head and pinnae. In tendency, and according to Hartung [55], the physical evaluation revealed smaller errors for frequency compared to time domain interpolation as well as for spline compared to linear interpolation, whereas original and minimum phase interpolation on average performed identical.

Difference thresholds that represent the minimally needed angular resolution of HATO were (a) estimated from perceptual evaluation, and (b) predicted based on the latter. Both reflect the superiority of torso interpolation: For the vast majority of tested algorithms, thresholds for torso interpolation outperform those of head interpolation by a factor of two to three. As assumed a priori, the median threshold of 10.5° for the noise stimulus and head interpolation (cf. Table 3.3) is comparable to results of Minnaar et al.[56]. For a source at $(315°; -30°)$, the authors found a resolution of 8° to be sufficient for interpolation artifacts to be inaudibly small. Note that the criterion of audibility applied by Minnaar et al. is stricter then the threshold criterion applied in our study which might account for the gap between the results. The similarity is due to the fact that in both cases HRTFs with different head-to-source orientations were used for interpolation.

While the perceptual evaluation was carried out using dynamic binaural synthesis allowing for head rotations in the horizontal plane, it can be assumed that different subjects listened to HRTFs for different head orientations during rating. This makes it likely that not all subjects discovered the head orientations where largest differences appeared. However, it seems unlikely that the dynamic auralization biased the results keeping in mind that (a) subjects were carefully instructed and trained to listen for differences at various head orientation, (b) an inspection of the raw data (hit rates in listening test I; thresholds in listening test II) suggest that most subjects actually detected differences, and (c) results are comparable to listening tests carried out using static binaural synthesis [57]. In turn, we suggest that the results are generalizable to a wide range of head orientations

[53] J. Blauert (**1997**). *Spatial Hearing. The psychophysics of human sound localization.*

[54] V. R. Algazi, et al. (**2001**a). "Elevation localization and head-related transfer function analysis at low frequencies" *J. Acoust. Soc. Am.*

[55] K. Hartung, et al. (**1999**). "Comparison of different methods for the interpolation of head-related transfer functions" in *16th Int. AES Conference.*

[56] P. Minnaar, et al. (**2005**). "Directional resolution of head-related transfer functions required in binaural synthesis" *J. Audio Eng. Soc.*

[57] P. Minnaar, et al. (**2005**). "Directional resolution of head-related transfer functions required in binaural synthesis" *J. Audio Eng. Soc.*

because different subjects evaluated HRTFs at different head orientations.

The interpolation of HATO requires HRTF data sets with a high resolution and various HATOs. Different proposals were made regarding the required resolution of source positions. Zhang et al. [58] transformed HRTFs into the spherical harmonic domain and found the reconstruction to be *reasonably accurate* if using 2304 HRTFs. Minnaar et al. [59] suggested that interpolation errors will remain inaudible for 1130 HRTFs if using minimum phase interpolation in the time domain. Consequently, a perceptual transparent representation of the HATO in the range of $\pm 75^\circ$ will require about 8,000 to 16,000 HRTFs using $\Delta\varphi_{\mathrm{HATO}} = 25^\circ$ (predicted 50% threshold percentile for torso interpolation; Table 3.4, #16). This appears to be feasible – even for human subjects – when considering fast HRTF measurement and modeling techniques [60,61].

[58] W. Zhang, et al. (**2012**). "On high-resolution head-related transfer function measurements: An efficient sampling scheme" *IEEE Transactions on Audio, Speech and Language Processing*.

[59] P. Minnaar, et al. (**2005**). "Directional resolution of head-related transfer functions required in binaural synthesis" *J. Audio Eng. Soc.*

[60] G. Enzner, et al. (**2013**). "Acquisition and representation of head-related transfer functions" in *The technology of binaural listening*, edited by J. Blauert.

[61] T. Huttunen, et al. (**2014**). "Rapid generation of personalized HRTFs" in *55th Int. AES. Conf.: Spatial Audio*.

3.6 Conclusion

In this study, we assessed the audibility of differences occurring during head rotations with constant or variable HATO, as well as the suitability of different algorithms for interpolating the HATO in HRTFs. To this end, we examined spectral and temporal deviations, and conducted two listening tests.

Although the effect of the torso on the HRTF is small compared to that of head and pinnae, we showed that differences between head rotations with constant and variable HATO were audible for the vast majority of source positions and audio contents. This suggests the importance of accounting for correct HATO at least if aiming at an authentic auralization, i.e. an auralization that is indistinguishable from a corresponding real sound field. This might, for example, be the case when benchmarking BRIRs obtained from numerical room modeling techniques against measured BRIRs.

Our evaluation of the interpolation of HATO in HRTFs showed that a grid width between 20° and 25° is sufficient when using torso interpolation, even for critical audio content and source positions. In this case, interpolation artifacts were below threshold for 50% of the subjects. A resolution of 8° was needed for artifacts to be subliminal for 95% of the subjects. If feasible, interpolation should be carried out in the frequency domain separately for the magnitude and unwrapped phase response.

This study was restricted to head rotations in the horizontal plane, because they were considered most important and critical. Nevertheless, future studies could also investigate the effect and interpolation of head rotations in elevation and roll. Moreover, it would be interesting to examine in how far interpolation algorithms in general – not only for HATO – can be applied to BRIRs, too, while assuming that reverberant sound fields will pose higher demands on interpolation algorithms. In addition, perceptual consequences of artifacts arising from larger interpolation intervals might be subjected to fur-

ther qualitative analysis, as this might be interesting for applications not demanding an authentic reproduction.

4
A high resolution and full-spherical head-related transfer function data base for different head-above-torso orientations

Fabian Brinkmann, Alexander Lindau, Stefan Weinzierl, Steven van de Par, Markus Müller-Trapet, Rob Opdam, and Michael Vorländer
(**2017**), *J. Audio Eng. Soc*, **65**(10), 841–848. DOI: 10.17743/jaes.2017.0033.
(Accepted manuscript. ©Audio Engineering Society)

Head-related transfer functions (HRTFs) were acoustically measured and numerically simulated for the FABIAN head and torso simulator on a full-spherical and high resolution sampling grid. Moreover, HRTFs were acquired for 11 horizontal head-above-torso orientations, covering the typical range of motion of $\pm 50°$, making it possible to account for head movements of the listeners in dynamic binaural auralizations in a physically correct manner. In lack of an external reference for HRTFs, measured and simulated data sets were cross-validated by applying auditory models for localization performance and spectral coloration, and by correlation analyses. The results indicate a high degree of similarity between the two data sets regarding all tested aspects, thus suggesting that they are free of systematic errors. The HRTF data base is publicly available from https://dx.doi.org/10.14279/depositonce-5718, and is accompanied by a wide range of headphone filters for use in binaural synthesis.

4.1 Introduction

Head-related transfer functions (HRTFs) capture the free field sound transmission from a sound source to the listeners ears. They incorporate all cues for sound localization such as interaural time and level differences (ITD, ILD) and spectral cues, that originate from scattering, diffraction, and reflection on the human pinnae, head, and body[1]. Using binaural synthesis and room acoustic simulation[2],

[1] H. Møller (**1992**). "Fundamentals of binaural technology" *Appl. Acoust.*

[2] M. Vorländer (**2008**). *Auralization. Fundamentals of acoustics, modelling, simulation, algorithms and acoustic virtual reality.*

HRTFs can thus be used to simulate spatial hearing, and open up a wide range of virtual auditory display applications such as guiding systems[3], game and mobile sound[4], or room acoustic design[5] and acoustic recreation of historic spaces[6].

Algazi *et al.*[7] showed that the torso effects the HRTF by means of reflecting or shadowing sound waves travelling towards the listeners ears. The reflection is strongest if source shoulder and ear are approximately aligned, and superimposes a comb filter to the HRTF with a magnitude of up to ± 5 dB. The first comb filter maxima occurs at approximately 700 Hz for sources at high elevations, and gradually increases in frequency with decreasing elevation. Shadowing occurs for sources well below the horizontal plane, and causes a high frequency damping of up to 25 dB that increases with decreasing sound source elevation. Perceptual investigations revealed that the cues induced by the torso and head are relevant for localizing the elevation of sources away from the sagittal median plane when pinna cues are absent[8]. Moreover, differences between head-above-torso orientations (HATOs) can be audible even for HRTFs that exhibit only weak torso effects[9]. Although the broadband interaural time and level differences (ITD, ILD) remain mainly unaffected by the HATO, it might be assumed that the HATO affects the ITD fine structure. This is known to be the case for the head, and was assumed to provide additional elevation cues and to help resolve front-back confusion[10]. Besides the influence of the torso on localization and timbre, dynamic HRTF cues related to head movements and HATO also affect other aspects of spatial hearing. It was for instance observed, that listeners naturally move their heads without moving the torso, when judging perceptual sound field qualities such as source width, or envelopment[11], and that head movements help to resolve front-back confusion and source elevation[12].

Nevertheless, currently available public HRTF sets – for an overview see[13] – were either measured for a fixed HATO or for dummy heads without torso. In the present study, we thus acquired HRTFs for multiple HATOs using acoustic measurements and numeric simulations as outlined in Section 4.2. In lack of an external reference for HRTFs, Section 4.3 details a cross-validation procedure that covers temporal and spectral aspects, as well as modeled localization performance. Please note that the current publication out-dates the preliminary post-processing of the acoustic measurements[14], and extends the initial cross-validation[15] to all HATOs and localization performance. Lastly, Section 4.4 describes the publicly available HRTF data base.

4.2 HRTF acquisition

HRTFs of the FABIAN head and torso simulator[16] were acquired for 11,950 source position with a dense spatial resolution (2° in elevation; 2° great circle distance in azimuth, cf. Figure 4.1A) that makes it suitable for a high order spherical harmonic representation. More-

[3] M. Bujacz, et al. (**2012**). "Naviton - a prototype mobility aid for auditory presentation of three-dimensional scenes to the visually impaired" *J. Audio Eng. Soc.*

[4] J. Sinker and J. Angus (**2015**). "Efficient compact representation of head related transfer functions for portable game audio" in *AES 56th International Conference: Audio for Games*.

[5] S. Pelzer, et al. (**2014**). "Integrating real-time room acoustics simulation into a cad modeling software to enhance the architectural design process" *Building Acoustics*.

[6] A. Pedrero, et al. (**2014**). "Virtual restoration of the sound of the hispanic rite" in *Forum Acusticum*.

[7] V. R. Algazi, et al. (**2001**a). "Elevation localization and head-related transfer function analysis at low frequencies" *J. Acoust. Soc. Am.*

[8] V. R. Algazi, et al. (**2001**a). "Elevation localization and head-related transfer function analysis at low frequencies" *J. Acoust. Soc. Am.*

[9] F. Brinkmann, et al. (**2015**b). "Audibility and interpolation of head-above-torso orientation in binaural technology" *IEEE J. Sel. Topics Signal Process.*

[10] V. Benichoux, et al. (**2016**). "On the variation of interaural time differences with frequency" *J. Acoust. Soc. Am.*

[11] C. Kim, et al. (**2013**). "Head movements made by listeners in experimental and real-life listening activities" *J. Audio Eng. Soc.*

[12] K. I. McAnally and R. L. Martin (**2014**). "Sound localization with head movement: implications for 3-d audio displays" *Frontiers in Neuroscience*.

[13] P. Majdak et al. (**2017**). *Collection of public head-related impulse response data bases* `http://sofacoustics.org/data/database/` (checked July 2017).

[14] F. Brinkmann, et al. (**2013**). "A high resolution head-related transfer function database including different orientations of head above the torso" in *AIA-DAGA 2013, International Conference on Acoustics*.

[15] F. Brinkmann, et al. (**2015**a). "Cross-validation of measured and modeld head-related transfer functions" in *Fortschritte der Akustik – DAGA 2015*.

[16] A. Lindau, et al. (**2007**). "Binaural resynthesis for comparative studies of acoustical environments" in *122th AES Convention, Convention Paper 7032*.

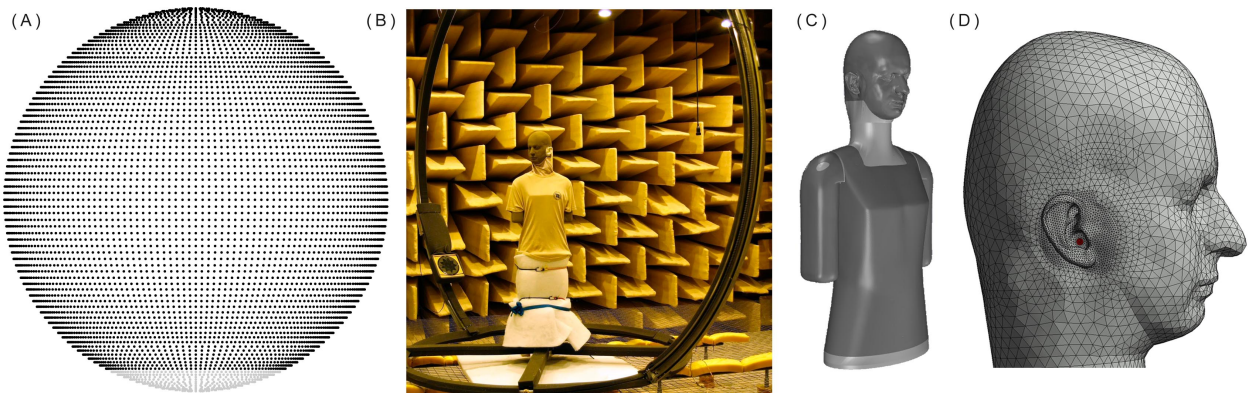


Figure 4.1: (A) Spherical sampling grid. Grey points show source positions below -64°. (B) Two arc source positioning system with FABIAN set up in its geometrical center. (C) 3D model of FABIAN. Light gray areas were manually inserted in post-processing. (D) Detail of the fine 3D surface mesh used for numerical simulation. Shaded area marks the microphone position.

over, these high resolution data sets were obtained individually for 11 HATOs which covered the typical range of head rotation of ±50° to the left and right[17], and with a resolution of 10° allowing for their perceptually transparent interpolation[18].

4.2.1 Acoustic measurements

Measurements were conducted in the anechoic chamber of the Carl von Ossietzky University Oldenburg ($V = 238$ m^3, cut-off frequency 50 Hz). To allow for subsequent HRTF identification, sine sweeps with a length of 2^{16} samples were recorded with DPA 4060 microphones at the entrances to FABIANs blocked ear canals (*recorded HRTF*), and at the position of the interaural center in the absence of FABIAN (*reference*). HRTFs were then derived by spectral division of the recorded HRTFs and the reference, yielding a signal to noise ratio (SNR) of 80 dB for ipsilateral and 55 dB for contralateral sources after averaging across four measurements. The sweep was designed in the frequency domain between 100 Hz and 20 kHz based on the group delay[19]. For achieving an almost frequency independent SNR, the energy of the sweep was set to be proportional to the background noise. The bandwidth was restricted according to the lower cut-off frequency of the loudspeakers used for measuring (Manger MSW bending-wave sound transducer in a custom made cubic closed box). *AKtools*[20] were used for sweep synthesis, deconvolution, as well as audio playback, and recording at a sampling rate of 44.1 kHz.

The two-arc-source-positioning system (TASP[21]), consisting of opposing semicircular arcs with a radius of 1.7 m, was used for positioning the sources with a precision of 0.1°. The two arcs could be rotated horizontally and were each equipped with a Manger MSW bending-wave sound transducer on vertically movable mounts (cf. Figure 4.1B). Due to mechanical restrictions, HRTFs could not be obtained for elevations below −64°. Before the measurements, FABIAN's interaural center was carefully aligned to the geometrical center of the TASP using a self-leveling Bosch PCL10 cross-line laser with the frontal viewing direction being defined by a laser pointer

[17] W. R. Thurlow, et al. (**1967**). "Head movements during sound localization" *J. Acoust. Soc. Am.*

[18] F. Brinkmann, et al. (**2015**b). "Audibility and interpolation of head-above-torso orientation in binaural technology" *IEEE J. Sel. Topics Signal Process.*

[19] S. Müller and P. Massarani (**2001**). "Transfer function measurement with sweeps. directors cut including previously unreleased material and some corrections" *J. Audio Eng. Soc.* (Original release).

[20] F. Brinkmann and S. Weinzierl (**2017**). "AKtools – An open software toolbox for signal acquisition, processing, and inspection in acoustics" in *142nd AES Convention, Convention e-Brief 309*.

[21] J. Otten (**2001**). "Factors influencing acoustical localization" Doctoral Thesis.

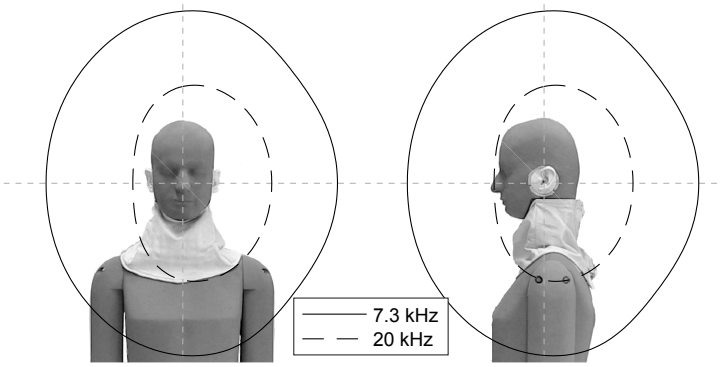


Figure 4.2: Directivity of one speaker from the TASP projected onto FABIAN by means of the -3 dB main lobe at 7.3 kHz and 20 kHz. Dashed crosses mark the position of the interaural center.

attached to FABIAN's neck joint.

Before measuring HRTFs for one HATO, 5.000 warm-ups of the sweep signal were played back through each loudspeaker to reduce their time variability from ± 4 dB to ± 1 dB. Nevertheless, a comparison of HRTFs for different HATOs showed differences of up to 2.5 dB within narrow frequency bands centered around approximately 0.5 and 2 kHz, that were consistent across source positions. These variations were attributed to long term changes in the loudspeakers frequency responses caused by loudspeaker aging and are visible in Figure 4.3 (bottom, left) by means of a horizontal line at 2 kHz for elevations between 14°, and 166°. The variations at 0.5 kHz are less distinct and not visible in Figure 4.3. Although this frequency range is also affected by the comb-filter caused by the shoulder reflection (U-shaped arches in Figure 4.3, bottom), the bandwidth of the observed artifacts is to narrow to be attributed to this effects.

By obtaining HRTFs through spectral division (deconvolution), the on-axis frequency response of the microphones, speakers, amplifiers, and converters cancels out, but the speakers directivity remains un-compensated. However, directivity measurements of the Manger transducers showed that FABIAN's shoulders and torso were within the -3 dB main lobe up to about 7 kHz. Consequently, the speakers directivities should be of negligible influence, because above 3-4 kHz, HRTFs are mainly influenced by the head and pinnae[22], which were within the main lobe up to 20 kHz (cf. Figure 4.2). The directivities were initially measured on 5x5 equal angle sampling grid and were comparable across speakers. However, the oval shape at high frequencies, which might be caused by the custom box design, was less pronounced for the second speaker. The main lobe widths were estimated in the spatially continuous spherical harmonics domain after applying a spherical harmonics transform of order 25[23].

Prior to deconvolution, a circular shift of 28 samples was applied to the recorded HRTFs to maintain causality and to ensure approximately 20 leading samples before the earliest peak in the head-related impulse responses (HRIRs). To discard reflections from opposing loudspeakers, HRIRs were truncated to 256 samples (5.8 ms). Finally, 10 (20) samples squared sine fade-ins (fade-outs) were applied.

[22] V. R. Algazi, et al. (**2001**a). "Elevation localization and head-related transfer function analysis at low frequencies" *J. Acoust. Soc. Am.*

[23] B. Rafaely (**2015**). *Fundamentals of spherical array processing*.

To obtain an estimate of the mechanical reliability of the TASP, four ever identical HRTFs – neutral head orientation, sources to the front, back, left and right – were measured before and after acquiring a set of HRTFs. Deviations in the HRTFs magnitude spectra within and between sets were generally below ± 1.5 dB, but increased to ± 10 dB at the contralateral ear, and in the vicinity of notches. This, however, is well comparable to previous dummy head HRTF measurements[24], while slightly larger variability was observed for human subjects[25].

[24] B. P. Bovbjerg, et al. (**2000**). "Measuring the head-related transfer functions of an artificial head with a high directional resolution" in *109th AES Convention, Preprint*

[25] H. Møller, et al. (**1995**b). "Head-related transfer functions of human subjects" *J. Audio Eng. Soc.*

4.2.2 Numeric simulations

Numeric HRTF simulation was done by means of the boundary element method (BEM) that requires a 3D surface mesh representation of FABIAN. This was generated in the following way: an initial point cloud representation was measured with a *GOM ATOS I* structured light scanner. A resolution of approximately $^1/_{100}$ mm for the head and pinnae, and $^1/_{10}$ mm for the torso was achieved by automatic alignment of overlapping scans, relying on manually inserted reference points and conducted with the surface matching algorithm of *ATOS Professional*. A non-uniform rational basis spline (NURBS) representation was built from the point cloud using *Geomagix Studio 12*. Subsequently, *Rhino 4* was used to (I) design a cylindrical neck with a seamless transition between head and torso, (II) to extend the torso bottom to its original size, (III) close screw holes in the arms, and (IV) connect the separate arm scans to the torso (cf. Figure 4.1C). At last, *Virtual.Lab Acoustics 13.1* was used for mesh generation, and calculating complex HRTF spectra at frequencies between 100 Hz and 22 kHz (100 Hz resolution). For acceleration, different triangular meshes were generated: a coarse mesh with edge lengths of 2 mm (pinnae), and 10 mm (head, and torso) was used for simulations up to 6 kHz, and a fine mesh with edge lengths of 2 mm (pinnae, and head), and 5 mm (torso) was used for the fast multipole method (FMM) BEM above 2 kHz (cf. Figure 4.1D). The chosen edge lengths fulfil the typical requirement of six elements per wavelength[26], and it was shown that a resolution of 1-2 mm sufficiently captures the details of the pinna geometry[27]. The overlapping region between 2 kHz and 6 kHz was used to verify that both simulations yielded identical results. Constant velocity boundary conditions were imposed to the mesh elements corresponding to the microphone at the entrances to the blocked ear canals. Otherwise, the mesh was assumed to be acoustically rigid. While this simplified assumption appears to be valid for human skin[28], it does not strictly hold for FABIAN's neck, which consists of a metal neck joint covered by a leather fabric, and was for simplicity modeled by a cylindrical shape that was smoothly extended towards the head and torso. Because the fabric of FABIAN's t-shirt with a thickness of less than 1 mm does not compare to existing experimental results for the effect of cloth in HRTF[29,30], it was also assumed to be rigid. Moreover, FABIAN's

[26] R. Ciskowski and C. Brebbia (**1991**). *Boundary Element Methods in Acoustics*.

[27] H. Ziegelwanger, et al. (**2015**b). "Numerical calculation of listener-specific head-related transfer functions and sound localization: Microphone model and mesh discretization" *J. Acoust. Soc. Am.*

[28] B. F. G. Katz (**2001**). "Boundary element method calculation of individual head-related transfer function. II. impedance effects and comparisons to real measurements" *J. Acoust. Soc. Am.*

[29] G. F. Kuhn (**1977**). "Model for the interaural time differences in the azimuthal plane" *J. Acoust. Soc. Am.*

[30] O. Kirkeby, et al. (**2007**). "Some effects of the torso on head-related transfer functions" in *122nd AES Convention, Convention Paper 7030*.

stand, which was wrapped in absorbing material during the acoustic measurements was not modeled due to computational restrictions.

As with the acoustic measurements, HRTFs were calculated by spectral division of the result at the sampling grid points by the analytical solution of a point source with the same volume velocity placed in the center of the coordinate system; the frequency bin at 0 Hz was set to 0 dB. HRIRs with a length of 441 samples, and 44.1 kHz sampling rate were obtained by inverse Fourier transform after mirroring the single sided spectra. Finally, the simulated HRIRs were windowed in the same way as their measured counterparts.

4.3 Cross-validation

A visual comparison of measured and simulated HRTFs showed a good agreement (cf. Figure 4.3). In lack of an external reference for HRTFs, cross-validation between measured and simulated data was already suggested by Turku *et al.*[31], who perceptually tested differences in localization and preference. Moreover, Jin *et al.*[32] assessed differences in head radii and spatial correlation, however, without providing evidence for the perceptual relevance of the suggested measures. In the current study, we physically conducted the cross-validation by comparing the temporal and spectral structure as well as the modeled median plane localization performance.

[31] J. Turku, et al. (**2008**). "Perceptual evaluation of numerically simulated head-related transfer functions" in *124th AES Convention, Preprint 7489*.

[32] C. Jin, et al. (**2014**). "Creating the sidney york morphological and acoustic recordings of ears database" *IEEE Trans. on Multimedia*.

4.3.1 Temporal structure

In theory, the time of arrival (TOA), i.e. the onset in the HRIRs, should be identical across measured and simulated data sets. However, average (and maximum) differences of $\tau = 1.2$ ($\tau = 4$) samples (28 μs and 91 μs) were observed between the two conditions, which equals a displacement of 9 mm (31 mm), or 0.3° (1°) ($c = 339$ m/s according to the average temperature during the measurements of 11.4° C, and the TASP radius of 1.7 m). Because the geometrical alignment of FABIAN was assumed to be close to perfect for the simulated HRIRs, differences in TOA can be caused by temperature fluctuations, and positioning inaccuracy during the acoustic measurements. The latter was supported by an analysis of τ across source positions, revealing slight discontinuities of up to about 3 samples (not shown here) that were attributed to the start and end points of the TASP rotation and the transition between the two loudspeakers. Moreover, observed temperatures fluctuations during the measurements of 3.1° C could induce an error of up to 1.2 samples (27 μs). The results of the TOA analysis suggest a high reliability of the setup, and that there should be no audible differences between measured and modeled HRIRs caused by mechanical inaccuracy or temperature fluctuation.

Nevertheless, the simulated data were used for correcting the TOA of the measured HRIRs, because time alignment was a prerequisite for the processing steps described in the next section. Alignment

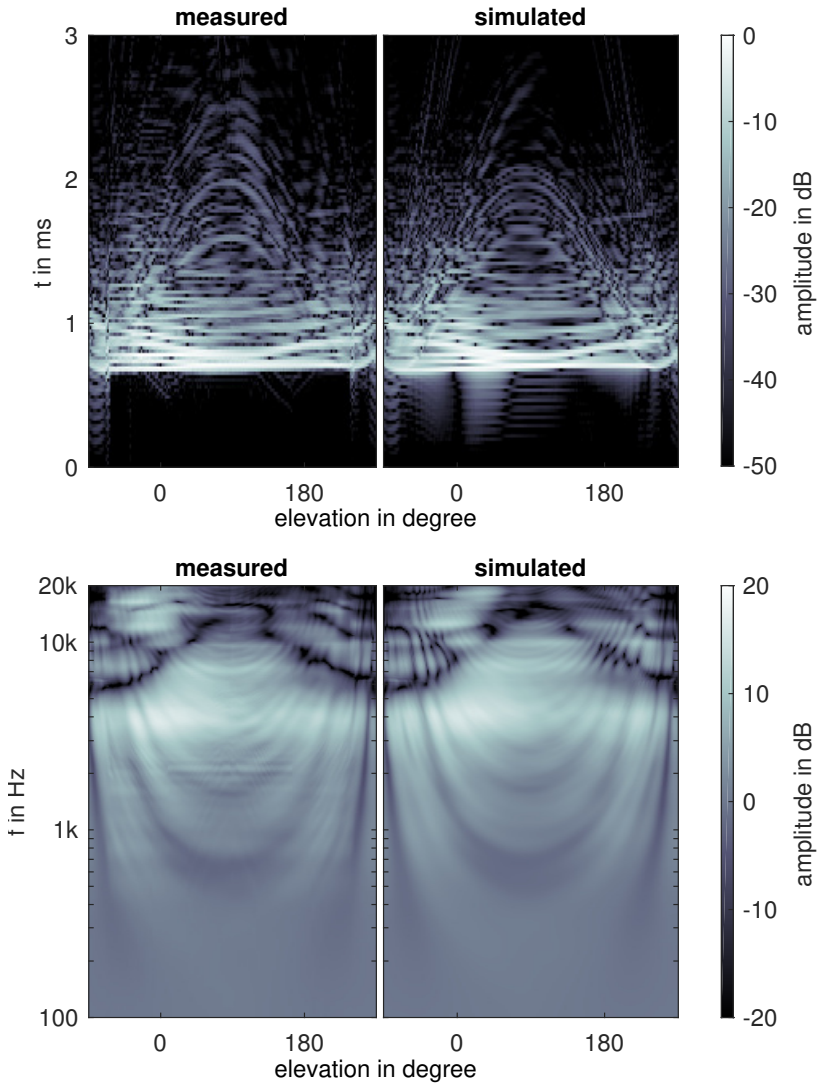


Figure 4.3: Comparison of left ear median plane HRIRs (top), and HRTFs (bottom) for neutral HATO. Elevation of 0°, and 180° denote sources to the front, and back of FABIAN, respectively.

was done using fractional delays[33] (Kaiser windowed sinc filters of order 70, 60 dB side lobe attenuation), with negligible magnitude and group delay distortions (< 0.1 dB; < 0.01 samples, $\forall f < 20$ kHz). As a result, the average cross-correlations between pairs of measured and simulated HRIRs increased from 0.51 to 0.93. Notably, the cross-correlation varied across HATO before the alignment (standard deviation $\sigma = 0.35$), showing the influence of the experimental setup, but was almost constant afterward ($\sigma = 0.05$). The fractional delaying induced changes in the broad band interaural time difference (ITD) of up to 66 ¯s for lateral sources (cf. Figure 4.4A). This however, is below average discrimination thresholds for broad band noise reported by Mossop and Culling[34], and was thus assumed to be perceptually irrelevant.

Besides differences in TOAs, simulated HRIRs exhibit more pre-ringing (cf. Figure 4.3, top right). Pre-ringing inevitably occurs in band limited digital signals if the TOA does not coincide with a sampling point of the impulse response. The fact that it is higher for the

[33] T. I. Laakso, et al. (**1996**). "Splitting the unit delay" *IEEE Signal Processing Magazine*.

[34] J. E. Mossop and J. F. Culling (**1998**). "Lateralization of large interaural delays" *J. Acoust. Soc. Am.*

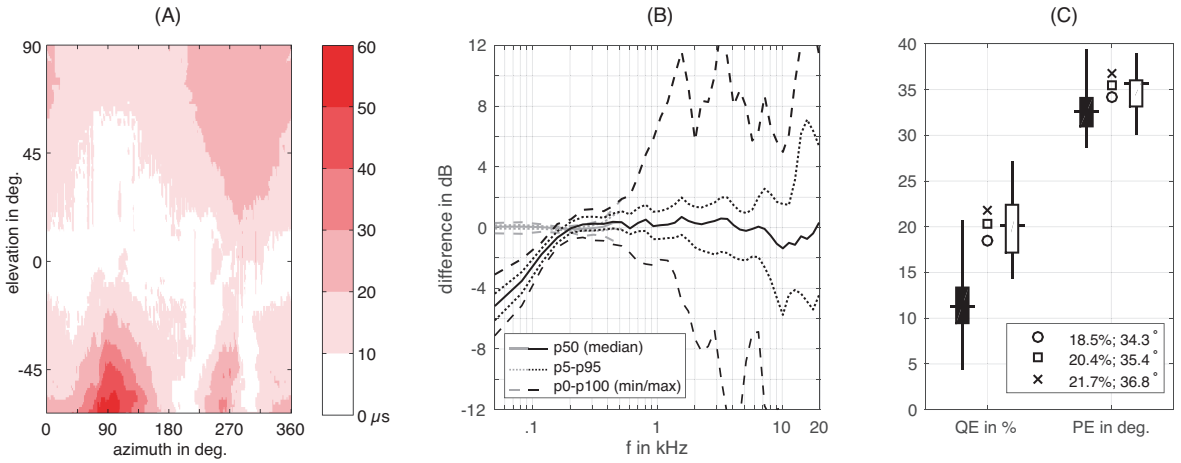


Figure 4.4: Cross-validation of measured and simulated HRTFs: (A) Maximal changes in broad band ITD across HATOs due to fractional delaying of measured HRIRs. (B) Spectral differences between measured and simulated HRTFs (averaged across source positions and HATOs) given by selected percentile ranges $p_{i,j}$ in 40 auditory filter bands. Black lines show differences before replacing measured data, gray lines show differences afterwards. (C) Modeled localization performance by means of PE and QE for measured (circles), simulated (squares), and measured vs. simulated (crosses) FABIAN HRTFs, accompanied by results for the ARI data base (black boxes), and a dummy head data base (white boxes). Horizontal lines show the median, boxes show the interquartile range, and vertical lines span between the minimum and maximum values.

simulated data, implies that HRIRs are less damped, which might be a consequence of assuming rigid boundary conditions during numeric simulation.

4.3.2 *Spectral structure*

Differences between measured and simulated HRTF magnitude spectra were analyzed in 40 auditory filter bands

$$\Delta G(f_c) = 10\log\frac{\int C(f,f_c)\,|\mathrm{HRTF}_{\mathrm{measured}}(f)|^2\,\mathrm{d}f}{\int C(f,f_c)\,|\mathrm{HRTF}_{\mathrm{simulated}}(f)|^2\,\mathrm{d}f}\,, \qquad (4.1)$$

where $C(f,f_c)$ are filters from the auditory modeling toolbox[35] with center frequency f_c, and 50 Hz $\leq f,f_c \leq$ 20 kHz. Results averaged across source positions and HATOs are shown as black lines in Figure 4.4B. Deviations below 200 Hz, where the level of the measured HRTFs is systematically lower, can be attributed to the band limited frequency response of the loudspeakers. Above 200 Hz, the median deviation p_{50} is smaller than $\pm$1 dB, indicating that both HRTF data sets are free of systematic errors. Moreover, 90% of the differences given by the $p_{5,95}$ percentile range are below $\pm$2 dB up to approximately 7 kHz, again emphasizing the very good agreement between measured and simulated data sets. The large difference between $p_{5,95}$, and $p_{0,100}$ in the range of 1-7 kHz indicates that $p_{0,100}$ is dominated by occasional outliers. Higher differences above 7 kHz of $\pm$10 dB and more originate from contralateral source positions where the HRIRs carry less energy, and mismatched HRTF notches across data sets. They might thus be perceptually less relevant at least for source localization, that is assumed to be dominated by the ipsilateral ear and the overall spectral shape across auditory filters[36].

Observed differences at high frequencies are difficult to correct, as they can either be caused by uncertainties in the measured HRTFs (e.g. positioning errors), or by simplifying assumptions of the BEM simulation (e.g. surface impedance). However, differences at low frequency can clearly be attributed to non-ideal loudspeaker charac-

[35] M. Slaney (**1998**). "Auditory toolbox. version 2" Technical Report #1998-010

[36] R. Baumgartner, et al. (**2014**). "Modeling sound-source localization in sagittal planes for human listeners" *J. Acoust. Soc. Am.*

teristics. Given the good fit of measured and simulated HRIRs at mid frequencies, it seems reasonable to assume that the simulated data can be used to replace the measured data at low frequencies. Consequently, measured and simulated HRTF magnitude and unwrapped phase spectra were combined separately using a linear fade between 200 and 500 Hz (gray lines in Figure 4.4B). Remaining differences below 200 Hz shown by $p_{0,100}$ are smaller than 0.5 dB and are caused by truncation and windowing.

The simulated data were also used to fill-up the missing HRTFs below -64° elevation that could not be measured due to mechanical restrictions. To obtain a smooth transition between the measured and modeled data, a linear fade was applied by interpolating magnitude and unwrapped phase spectra for elevations between -60° to -64°. This caused a slight discontinuity in the HRIRs (vertical line in Figure 4.3, top, left) but was preferred to spherical harmonics based extrapolation[37] that resulted in low-passed HRTFs at the missing positions due to a low truncation order.

[37] J. Ahrens, et al. (**2012**). "HRTF magnitude modeling using a non-regularized least-squares fit of spherical harmonics coefficients on incomplete data" in *APSIPA Annual Summit and Conference*.

4.3.3 *Median sagittal plane localization*

The localization performance in the median sagittal plane was assessed using the probabilistic model of Baumgartner *et al.*[38], which compares the spectral structure of a target HRTF set to a set of template HRTFs. Based on this comparison, it estimates quadrant errors (QE) – specifying the percentage of front-back/up-down confusion – and polar errors (PE) – specifying the angular rms error between actual and modeled source positions. Correspondingly, if target and template HRTFs are identical, the model indirectly assesses the uniqueness of an HRTF within the data set compared to the remainder, whereas, if passing different HRTF sets, it assesses the similarity between them. To be comparable to Baumgartner *et al.*, we assumed a median listener sensitivity of $S_l = 0.76$ and considered only elevations above $-30°$.

[38] R. Baumgartner, et al. (**2014**). "Modeling sound-source localization in sagittal planes for human listeners" *J. Acoust. Soc. Am.*

The within localization performance averaged across HATOs exhibits a very good agreement between data sets: QEs for simulated HRIRs (squares in Figure 4.4C) are only 2% worse than those of their measured counterparts (circles in Figure 4.4C), and PEs differ by only 1°. For external validation, localization performance was also modeled using HRTFs of all human subjects from the ARI data base (solid lines in Figure 4.4C), and all dummy heads (dashed lines in Figure 4.4C) available from[39]. Average localization performance is better for human subjects than for dummy heads, a fact that deserves further investigation, however, the estimates for FABIAN are in excellent agreement with the dummy head results. Since the localization model relies on spectral structure, this result indicates a general validity of the FABIAN HRIRs. In addition, the between localization performance, i.e. passing measured HRTFs as template and simulated HRTFs as target (crosses in Figure 4.4C), is only marginally worse than the within performance. This supports the hypothesis

[39] P. Majdak et al. (**2017**). *Collection of public head-related impulse response data bases* `http://sofacoustics.org/data/database/` (checked July 2017).

that large spectral differences given by $p_{0,100}$ in Figure 4.4B are perceptually less relevant regarding localization.

4.4 Database

The FABIAN head and torso simulator data base is publicly available from `https://dx.doi.org/10.14279/depositonce-5718`. Measured and simulated head-related impulse responses (HRIRs) are given in the SOFA format[40]. They are accompanied by minimum phase common transfer functions (CTFs) and their inverse. Based on the assumption of a diffuse sound field, CTFs were computed by power averaging HRTF magnitude spectra

$$\mathrm{CTF}(f) = \sqrt{\sum_i |HRTF_i(f)|^2 \, w_i} \,, \tag{4.2}$$

with frequency f and averaging weights w_i. The latter were computed according to the areas of spherical rectangles around each source position in Figure 4.1A, normalized to $\sum_i w_i = 1$. The CTFs were calculated separately for measured and simulated HRTF sets, and averaged across HATOs.

For convenience HRTF data sets were transformed into spherical harmonics (SH) coefficients, separately for each HATO by applying the 35th order discrete spherical harmonics transform (DSHT) to the complex HRTF spectra[41]. This converts the spatially discrete HRIR data sets into a continuous representation, and reduces the the amount of data by a factor of about 10 (256 real HRIR values $\times$ 119050 source positions vs. 129 frequencies $\times$ $(35+1)^2$ complex SH coef.). HRIR interpolation between source positions (in the SH domain) and HATOs (in the frequency domain[42]) can be done using `AKhrirInterpolation` from the *AKtools*[43].

To foster the use of the data base, headphone impulse responses (HpIRs) measured on FABIAN, and corresponsding inverse filters of about 35 models including widespread Sennheiser HD600, AKG K701, and Beyerdynamic DT770 headphones are part of the data base. The filters were calculated by means of the regulated least mean square inversion[44]. Parametric equalizers (PEQs) – fitted to the notches in the headphone transfer functions – were used for regularization to avoid an exact inversion in these cases. They are intended for headphone equalization in binaural synthesis. In this context, the inverse CTFs could be used as a generic headphone compensation filter considering the diffuse field HRTF to be a common target curve in headphone development[45]. Additionally, all generated 3D meshes are included, as well as example auralizations of fixed and moving sources.

4.5 Summary

An HRTF data set of the FABIAN head-and-torso simulator was generated by measuring and simulating HRTFs for a high resolution, full

[40] AES Standards Comittee (**2015**). *AES69-2015: AES standard for file exchange - Spatial acoustic data file format*.

[41] Eq. (1.9), and (3.34) in B. Rafaely (**2015**). *Fundamentals of spherical array processing*.

[42] F. Brinkmann, et al. (**2015**b). "Audibility and interpolation of head-above-torso orientation in binaural technology" *IEEE J. Sel. Topics Signal Process.*

[43] F. Brinkmann and S. Weinzierl (**2017**). "AKtools – An open software toolbox for signal acquisition, processing, and inspection in acoustics" in *142nd AES Convention, Convention e-Brief 309*.

[44] A. Lindau and F. Brinkmann (**2012**). "Perceptual evaluation of headphone compensation in binaural synthesis based on non-individual recordings" *J. Audio Eng. Soc.*

[45] H. Møller, et al. (**1995**a). "Transfer characteristics of headphones measured on human ears" *J. Audio Eng. Soc.*

spherical sampling grid, and eleven head-above-torso orientations. A detailed cross-validation showed a very good agreement in terms of temporal and spectral structure, as well as modeled localization performance. In turn, the simulated HRTFs were used to correct the time of arrivals and low frequency response in their measured counterparts. The data set is publicly available, and comprises 11,950 HRIRs for each HATO, corresponding spherical harmonics coefficients, 3D surface meshes, numerous headphone filters for binaural synthesis, and auralizations of fixed and moving sources for comparing measured and simulated HRIRs[46]. A perceptually transparent interpolation between different HATOs can be done using *AKtools*[47]. Future work could evaluate the influence of the head-above-torso orientation on the ITD fine structure.

[46] F. Brinkmann, et al. (**2017**b). *The FABIAN head-related transfer function data base* `https://dx.doi.org/10.14279/depositonce-5718.2`.

[47] F. Brinkmann and S. Weinzierl (**2017**). "AKtools – An open software toolbox for signal acquisition, processing, and inspection in acoustics" in *142nd AES Convention, Convention e-Brief 309*.

5
A benchmark for room acoustical simulation. Concept and database

Fabian Brinkmann, Lukas Aspöck, David Ackermann, Rob Opdam, Michael Vorländer, and Stefan Weinzierl. *Acta Acust. united Ac.*
(Under review. CC-BY 4.0)

Room acoustical simulations are usually evaluated by comparing them to corresponding measurements as a benchmark. However, it proved to be challenging to provide a precise representation of the room geometry, the source and receiver characteristics, and the absorption and scattering coefficients to be re-modeled in the simulation. We aim to overcome this shortcoming by providing a database that can serve as a Benchmark for Room Acoustical Simulations (BRAS) and is permanently available for researchers and developers of room simulation software. The database includes a selection of acoustic scenes such as "single reflection", or "diffraction around an infinite wedge" which isolate specific acoustic phenomena, as well as four complex "real-world" environments of different sizes. This article introduces the concept of the BRAS along with the description of the acoustic scenes, the acquisition of impulse responses for omni-directional and binaural receivers, and the identification of the boundary conditions. In addition, the implication of measurement errors are discussed, and possible evaluation schemes and methods are introduced. The BRAS is publicly available from `https://dx.doi.org/10.14279/depositonce-6726.2`. The free license under which it is provided allows for future extensions such as additional scenes or improved data due to advanced measurement techniques.

5.1 Introduction

Room acoustical simulation enables the numerical calculation of sound propagation in enclosed and open spaces. Corresponding algorithms are either based on the assumptions of geometrical acoustics (GA), considering sound to propagate as rays, or on numerical solution of the wave equation, applying different techniques such as finite-difference methods (FDM), the finite element method (FEM), or the

boundary element method (BEM)[1]. Due to the high computational effort, the latter are, however, mainly applied for low frequencies and simple geometries so far.

Room acoustical simulations have a broad field of application including the acoustical reconstruction of historic venues[2,3], the design of new concert hallsauto[4], classrooms, open offices or train stations and stadiums[5], the planning of urban areas[6], the creation of complex game audio scenarios[7], the investigation of particular room acoustic phenomena[8,9] or the experimental study of the impact of room acoustics on speech perception[10] and musical performance[11], to name just a few recent examples. Many of these applications make use of the possibility to listen through the virtual ears of a dummy head or head-and-torso simulator – a process which was coined auralization[12]. At the same time, there is no undivided confidence in the accuracy of room acoustical simulations, when it comes, for example, to the design of new performance venues for music and speech, where acoustic scale models are still an important alternative with specific advantages[13]. The multitude of applications and the importance of acoustical simulation thus necessitates a comprehensive evaluation of the corresponding algorithms, especially if considering that all of them have underlying simplifying assumptions or a limited frequency range of operation (for an overview see[14]).

Evaluations of room acoustical simulation algorithms with this purpose were carried out in the three international round robins on room acoustical computer simulation, termed RR-I to RR-III in the following[15,16,17,18]. In these round robins, different information was provided to the participants at different phases. In phase I of RR-I and RR-II, the participants had to estimate the geometry and the boundary conditions themselves from architectural plans and written information ("3 mm carpet"); in phase II the data was harmonized based on a common 3D model and boundary conditions estimated by room acoustical measurements. In RR-III, absorption and scattering coefficients for one wall and the ceiling of the room were measured in the reverberation room, and taken from tabulated data otherwise.

Two databases with analytically defined test scenarios were established by Otsuru *et al.*[19,20] and Hornikx *et al.*[21]. They are intended for cross-validation of wave based simulation algorithms, and provide perfect scene descriptions without providing measured or analytical references.

The two examples demonstrate that any evaluation of room acoustical simulation software has to define a strategy how to provide a suitable reference for the simulation, and how to control the uncertainties related to this reference. Both aspects have consequences for the content of the presented database.

A first source of uncertainty is the geometric model of the acoustic scene. In most geometric room acoustical simulation software it is favorable to dispense with the representation of small surface structures below 0.5 m[22]. For wave-based simulations, on the other

[1] M. Vorländer (**2008**). *Auralization. Fundamentals of acoustics, modelling, simulation, algorithms and acoustic virtual reality*.

[2] J. H. Rindel (**2011**b). "The ERATO project and its contribution to our understanding of the acoustics of ancient theatres" in *The Acoustics of Ancient Theatres Conference*.

[3] S. Weinzierl, et al. (**2015**). "The acoustics of Renaissance theatres in Italy" *Acta Acust. united Ac.*

[4] H. Kuttruff (**1991**). "Digital simulation of concert hall acoustics and its applications" *Acoustic Bulletin*.

[5] P. Chevret (**2015**). "Advantage of the incoherent uniform theory of diffraction for acoustic calculations in open-plan offices" *J. Acoust. Soc. Am.*

[6] M. Hornikx (**2016**). "Ten questions concerning computational urban acoustics" *Building and Environment*.

[7] C. Schissler, et al. (**2016**). "Efficient hrtf-based spatial audio for area and volumetric sources" *IEEE Transactions on Visualization and Computer Graphics*.

[8] L. Shtrepi, et al. (**2015**). "Objective and perceptual assessment of the scattered sound field in a simulated concert hall" *J. Acoust. Soc. of Am.*

[9] S. Lu, et al. (**2016**). "The influence of shape design on the acoustic performance of concert halls from the viewpoint of acoustic potential of shapes" *Acta Acusts. united Ac.*

[10] M. Cipriano, et al. (**2017**). "Combined effect of noise and room acoustics on vocal effort in simulated classrooms" *J. Acoust. Soc. Am.*

[11] Z. S. Kalkandjiev and S. Weinzierl (**2015**). "The influence of room acoustics on solo music performances: An empirical investigation" *Psychomusicology: Music, Mind, and Brain*.

[12] M. Kleiner, et al. (**1993**). "Auralization – An overview" *J. Audio Eng. Soc.*

[13] J. H. Rindel (**2011**a). "Room acoustic modelling techniques: A comparison of a scale model and a computer model for a new opera theatre" *Building Acoustics*.

[14] L. Savioja and U. P. Svensson (**2015**). "Overview of geometrical room acoustic modeling techniques" *J. Acoust. Soc. Am.*

[15] M. Vorländer (**1995**). "International round robin on room acoustical computer simulations" in *15th International Congress on Acoustics*.

[16] I. Bork (**2000**). "A comparison of room simulation software – The 2nd round robin on room acoustical computer simulation" *Acta Acust. united Ac.*

[17] I. Bork (**2005**a). "Report on the 3rd round robin on room acoustical computer simulation - Part I: Measurements" *Acta Acust. united Ac.*

hand, a precise model is desirable, and even for algorithms based on ray tracing, the exact threshold of resolution may depend on the way scattering and diffraction is treated. For both approaches, the desired resolution may also depend on the considered frequency band. Therefore, the way a primary acoustic structure is modified and possibly simplified for simulations should, according to the authors, be considered as part of the simulation itself. This is true for the meshing methods for finite element simulations as well as for the geometric simplification for ray simulations. It should not be anticipated by manipulating the reference data in a way that would necessarily favour certain algorithms and certain frequencies over others. Therefore, the BRAS provides scene geometries above the resolutions commonly used in GA including structures below 50 cm.

A second source of uncertainty are the boundary conditions. In RR-I to RR-III and in similar investigations[23,24], the lack of valid boundary conditions data was identified as one of the most important factors why room acoustical simulations differ from measured results. For complex rooms such as concert venues or lecture halls, a comprehensive specification of absorption and scattering for all boundaries is practically impossible, because neither can all different surfaces with their different types of installation be measured in the laboratory, nor are any (standardized) full-range measurement techniques available to determine them in situ. Fitting the input parameters according to measurements of the reverberation time, on the other hand, may be a pragmatic solution for many problems in room acoustics planning. As a procedure for the evaluation of room acoustical simulation algorithms, however, it would contain an element of circular reasoning. If both the premises (the boundary conditions) and the success of the simulation are determined by the same measurement (of room acoustical parameters), the test will always tend to confirm the quality of the simulation algorithm. For these reasons, the boundary conditions provided within the BRAS were directly measured or deduced from measurements whenever possible.

A third source of uncertainty is the behaviour of the sources and receivers which are an integral part of the acoustic transfer function. In RR-I to RR-III, the reference measurements were done with industry-standard dodecahedron loudspeakers, whereas for the simulations, perfect omnidirectional sources were assumed. It was, however, shown that the non-ideal directivity of standard dodecahedron loudspeakers, even if they are compatible with the requirements according to ISO 3382, can be observed even at late parts of measured RIRs[25], and causes a measurement uncertainty above the JND for different room acoustical parameters and frequencies above 500 Hz[26]. To allow for an accurate analysis, the BRAS therefore provides measured directivities in high spatial resolution for all sound sources and the binaural receiver used.

The database presented here contains a collection of 11 acoustical scenes, each of which highlights certain acoustical phenomena and

[18] I. Bork (**2005**b). "Report on the 3rd round robin on room acoustical computer simulation - Part II: Calculations" *Acta Acust. united Ac.*

[19] T. Otsuru, et al. (**2005**). "Constructing a database of computational methods for environmental acoustics" *Acoust. Sci. & Tech.* , available from `http://news-sv.aij.or.jp/kankyo/s26/AIJ-BPCA/A0-1F/index.html` (Assessed: Mar. 2019).

[20] T. Sakuma, et al. (**2002**). "A round-robin test programm on wave-based computational methods for room-acoustic analysis" in *Forum Acusticum.*

[21] M. Hornikx, et al. (**2015**). "A platform for benchmark cases in computational acoustics" *Acta Acust. united Ac.* , available from `https://eaa-bench.mec.tuwien.ac.at/main/` (Assessed: Mar. 2019).

[22] M. Vorländer (**2008**). *Auralization. Fundamentals of acoustics, modelling, simulation, algorithms and acoustic virtual reality* , p. 176.

[23] S. Pelzer, et al. (**2011**). "Quality assessment of room acoustic simulation tools by comparing binaural measurements and simulations in an optimized test scenario (a)" *Acta Acust. united Ac.*

[24] L. Aspöck, et al. (**2016**). "Acquisition of boundary conditions for a room acoustic simulation comparison" in *ISMRA.*

[25] T. Knüttel, et al. (**2013**). "Influence of "omnidirectional" loudspeaker directivity on measured room impulse responses" *J. Acoust. Soc. Am.*

[26] R. S. Martín, et al. (**2007**). "Influence of the source orientation on the measurement of acoustic parameters" *Acta Acust. united Ac.*

certain spatial configurations, so that they can be used as a reference to evaluate the ability of room acoustical simulation software to model these phenomena and configurations. For a number of relatively simple yet practically relevant geometries the required input data can be given with reliable accuracy (scenes 1–8). In addition, the database contains three complex environments of different size (scenes 9–11), for which parts of the input data can only be approximated. These are nevertheless practically relevant in order to test in how far real-world scenarios such as lecture halls and concert venues can be modelled based on input data with a practically achievable precision.

For each of the scenes, a number of impulse responses for omnidirectional and binaural receivers has been determined by measurements, so that the quality of the corresponding simulations can later be evaluated both in the physical and in the perceptual domain. For the latter, the BRAS database contains head-related transfer functions (HRTFs) corresponding to the same head and torso simulator that was used for the binaural measurements.

The article introduces the eleven acoustic scenes of the BRAS (Section 5.2), details the acquisition of the input data (Section 5.3), and gives a brief overview of the database organization (Section 5.4). A discussion of the implication of measurement uncertainties is given in Section 5.5. More details, such as the exact source and receiver positions, the data formats/structure, and pictures of the scenes and the included acoustic materials are contained in the documentation of the database itself[27].

[27] L. Aspöck, et al. (**2019**). *BRAS – A Benchmark for Room Acoustical Simulation* https://dx.doi.org/10.14279/depositonce-6726.2.

5.2 *Acoustic Scenes*

An overview of the eleven scenes along with the number of contained source and receiver positions is given in Tab. 5.1 and Fig. 5.1. For a detailed technical description of each scene, including the exact geometry and source/receiver positions, please refer to the BRAS documentation[28].

[28] L. Aspöck, et al. (**2019**). *BRAS – A Benchmark for Room Acoustical Simulation* https://dx.doi.org/10.14279/depositonce-6726.2.

Scene 1 features a single reflection on a quasi infinite rigid, absorbing, and scattering surface for different angles of sound incidence. A reflection on and the diffraction around rigid and absorbing finite plates of two sizes was measured in scene 2. Impulse responses were acquired for different angles of sound incidence, and receiver positions in front of and behind the plate. Despite its geometric simplicity, this scene is challenging as diffraction and sound transmission around and through the plate has to be modeled, either with extended geometrical or wave based methods.

Scenes 3 – 8 aim at recreating simplified versions of relevant real life scenarios: The reflection between parallel plates (scene 3) evokes a flutter echo that is often problematic in room acoustics. Reflector arrays (scene 4) are frequently used in concert halls to direct early reflections to the audience area. Diffraction around wedges and bodies (scenes 5 & 6) are relevant in noise mapping and urban acous-

Figure 5.1: The eleven scenes included in the BRAS.

#	Name	Date	RIR	BRIR
1	single reflection (infinite plate)	Nov. 2015	3/3	1/1
2	single reflection & diffraction (finite plate)	Dec. 2015	6/5	–
3	multiple reflection (parallel finite plates)	Dec. 2015	1/1	1/1
4	single reflection (reflector array)	Dec. 2015	6/6	–
5	diffraction (infinite wedge)	Nov. 2015	4/4	1/1
6	diffraction (finite body)	Dec. 2016	3/3	–
7	multiple diffraction (seat dip effect)	Dec. 2016	2/4	–
8	coupled rooms (office & reverb. chamber)	Nov. 2015	2/2	2/2
9	small room (seminar room)	Nov. 2015	2/5	5/1
10	medium room (chamber music hall)	Dec. 2015	3/5	5/1
11	large room (auditorium)	Dec. 2015	2/5	5/1

Table 5.1: Overview of the BRAS scenes. Columns RIR and BRIR give the number of source/receiver positions used for measuring impulse responses with omnidirectional (RIR) and binaural (BRIR) receivers. The solid horizontal line separates the simple scenes 1 – 8 from the complex scenes 9 – 11.

tics, and diffraction on a repeated structure caused by grazing sound incidence (scene 7) occurs in audience areas and is well known for causing the seat-dip effect. Coupled volumes (scene 8) are often used to achieve variable acoustics in concert halls, or can be found behind theatre stages. While the input data for scene 8 has to be partly estimated, it is sufficiently accurate for comparing the energy decay of coupled rooms if the estimates are fitted to measured quantities.

Scenes 9, 10 and 11 are real-world rooms of different size that exhibit relatively simple shapes on one hand, but contain aspects that are challenging to model on the other: The empty seminar room with a volume of 144 m^3 and the concert hall with a total volume of 3,320 m^3 can be described as shoebox rooms. However, the seminar room has two plastered light weight walls where sound transmission and resonance have an effect on the absorption for frequency bands below 125 Hz[29], and the concert hall has 982 m^3 coupled volumes behind the stage and above the hall. The auditorium (scene 11, $V \approx 8,650\,m^3$) has a slightly more complex shape, including ceiling reflector arrays.

While all of these scenarios are known to be a challenge for room acoustical simulation based on GA, wave based methods can in theory well simulate all of them. Here the challenges lie for instance in the implementation of directional sources and receivers, the computational demands of broad-band simulations, or coupling wave based to GA simulations[30].

[29] For the smaller light weight wall the resonance frequency was measured between 45 Hz and 54 Hz.

[30] B. Hamilton (**2016**). "Finite difference and finite volume methods for wave-based modelling of room acoustics" Doctoral Thesis.

For all scenes, room impulse responses (RIRs) were measured with a 2-way active Genelec 8020c studio monitor, and either with half inch G.R.A.S. or Bruel&Kjær (B&K) microphones. To provide room acoustical parameters in accordance with ISO 3382-1[31], a additional custom made dodecahedral speaker was used in scenes 8 – 11. Binaural room impulse responses (BRIRs) were measured in scenes 9 – 11 using a 2-way active QSC-K8 PA speaker with positions imitating a string quartet with a singer. In scenes 1, 3, 5 and 8 BRIRs were measured at selected positions using the Genelec 8020c loudspeaker. BRIRs were measured for head rotations to the left and right of $\pm 44^\circ$ in steps of 2°.

Scenes 1 and 5 – 9 were measured at the hemi anechoic chamber, the reverberation chamber, and the seminar room at RWTH Aachen University. Scenes 2 – 4 and 11 were measured in the anechoic chamber and the Audimax at TU Berlin. Scene 10 was measured at the chamber music hall of Konzerthaus Berlin.

[31] ISO 3382-1 (**2009**). *Measurement of room acoustic parameters – Part 1: Performance spaces*.

5.3 *Data acquisition*

All acoustic measurements were conducted with a sampling rate of 44.1 kHz, and all impulse responses were obtained by swept sine measurements and spectral deconvolution[32]. All measurements were supervised and processed by the three primary authors. To assure consistency across the data of different scenes, a standardized measurement protocol, identical equipment, as well as identical measurement and post-processing scripts were used. The only differences among measurement and post-processing of different scenes were the length of the sine sweeps and the final impulse responses, which were both adjusted to the level of reverberation and background noise.

[32] S. Müller and P. Massarani (**2001**). "Transfer function measurement with sweeps. directors cut including previously unreleased material and some corrections" *J. Audio Eng. Soc.* (Original release).

5.3.1 *Scene Geometry*

Generating the scene geometry can be split in two parts: The acquisition of the room geometry, and positioning the objects inside the room. The latter was done relative to a predefined reference point with the help of self leveling cross line lasers (Bosch Quigo, precision ± 0.8 mm/m), a laser distance meter (Bosch DLE 50 Professional, precision ± 1.5 mm), and a laser angle measurer (geo-FENNEL EL 823, equipped with a Hama LP-21 laser pointer, precision $\pm 0.5^\circ$). The reference points for positioning the sources and receivers are listed in Table 5.3.1. They are identical to the center of rotation during the directivity measurements (cf. Section 5.3.3 and 5.3.4). In scenes 2 - 4, the objects had to be placed on the wire-woven floor of the fully anechoic chamber. Because the floor's inclination slightly changed due to the weight of the objects, the positions were iteratively adjusted with the help of a plump bob and a remote camera until the objects were perpendicular to the floor. Moreover, the weight of the objects was distributed to a larger area using stands (scenes 2 & 3),

Device	Reference point
Genelec 8020c (loudspeaker)	Point on the front panel midway between the outmost position of the low/mid frequency driver and the tweeter
QSC-K8 (loudspeaker)	Point on the protective grid in front of the tweeter
Dodecahedron (loudspeaker)	Low-frequency unit: center of circular opening; Mid-frequency unit: center of sphere; High-frequency unit: center of sphere
B&K type 4134 G.R.A.S. 40AF (microphones)	Point on the center of the protective grid in front of the membrane
FABIAN (dummy head)	Interaural center, defined as the midpoint of the line that connects the entrances of the two ear channels

Table 5.2: Definition of the reference points that define the positions of the sources and receivers. The orientation of the transducers is documented in the BRAS separately for each scene.

or objects were hung from the ceiling (scene 4). The precision of placing the objects within the scene was verified by estimating the direct sound arrival time from the ten times up-sampled measured impulse responses (IRs) by means of threshold based onset detection[33]. A comparison of the acoustically estimated arrival times to the geometrical values given in the scene descriptions showed absolute differences up to 5.5 cm (3.4 cm on average) for scenes 2 – 4, and 2.3 cm (1.6 cm on average) for scenes 1 and 5 – 7. For a source at 3 m distance, a positioning inaccuracy of 5.5 cm would cause negligible magnitude errors of maximally $20\log_{10}(3.055\text{ m}/3\text{ m}) = 0.16$ dB between the intended and actual sound pressure level, and an angular displacements of maximally $\arctan(0.055\text{ m}/3\text{ m}) = 1^\circ$ between the intended and actual source/receiver positions under the assumption of a receiver in the far field of a point-like source. Considering the Genelec 8020c speaker with a woofer-to-tweeter distance of 11 cm, and the smallest source-to-receiver distance of 3 m, the assumptions above appear to be valid (cf. Section 5.3.4). The additional check of the positioning precision by means of onset detection could not be done for scenes 8 – 11, because i) the Genelec loudspeaker was not facing the receivers, and ii) the dodecahedron source consists of multiple drivers, which substantially widens the direct sound impulse and caused errors in the onset detection. However, the positioning was done in analogy to scenes 1 & 5 - 7 using lasers, and the floor was planar which made it easy to measure the source-to-receiver distances. It is thus reasonable to assume that the positioning accuracy is comparable to that of scenes 1 and 5 – 7.

For scenes 8 – 10, the room geometry was acquired by manually measuring the positions of corners and important points with a TOPCON EM-30 laser distance meter (precision ± 3 mm) mounted on a VariSphear scanning microphone array[34]. The laser distance meter could be rotated in azimuth and elevation using two computer con-

[33] A. Andreopoulou and B. F. G. Katz (**2017**). "Identification of perceptually relevant methods of inter-aural time difference estimation" *J. Acoust. Soc. Am.*

[34] B. Bernschütz (**2013**). "A spherical far field HRIR/HRTF compilation of the Neumann KU 100" in *AIA-DAGA 2013, International Conference on Acoustics*.

trollable motors (Schunk PR070, minimum step width 0.001°). For scene 9, all points could be scanned from a single position, while some points were blocked in the remaining scenes. In these cases, the rooms were scanned from different positions, and duplicate points were used to align the point clouds. Euclidean distances between duplicate points were 7 mm on average. A check for planarity of the major room surfaces (walls, floor, ceiling) was done by fitting a plane through the points of each surface. Mean absolute deviations in direction of the surface normal of 2.6 cm were corrected by moving the points on to the surface. Maximum deviations of 6 cm and 8.5 cm occurred in two cases inside the reverberation chamber of scene 8 where the corner points were not clearly defined and for a wall that showed small irregularities. Since the reverberation chamber is solely intended as the coupled volume for the laboratory room of scene 8, this error can considered to be negligible. In addition the RMS Hausdorff distance between the non-planar and planar surfaces was calculated using METRO[35]. The average across rooms of 7.8 mm suggests, that violations of planarity were small in general. Afterwards, the 3D modeling software SketchUp was used to design the final room model based on the post-processed point clouds. The volume above the ceiling of scene 10 was measured by hand using the Bosch DLE 50 laser distance meter.

[35] P. Cignoni, et al. (**1998**). "Metro: Measuring error on simplified surfaces" *Computer Graphics Forum*.

Due to the complexity of scene 11, the initial room model was designed based on architectural drawings, and validated against a set of 15 manually measured distances between relevant points in the room (e.g. the height width and depth at different positions in the room). Deviations between measured and initially modeled distances were 20 cm on average (SD =16 cm), and the model was corrected according to the measurements by adjusting the position of the floor, ceiling and walls. In a final step, material names were assigned to each surface in the room models using SketchUp, and the positions and orientations of the sources/receivers were inserted.

The geometry acquisition aimed at obtaining a high level of detail rather than simplifying parts of the geometry, which should be done prior to running an acoustical simulation. The models therefore also include structures below 50 cm that might have to be simplified depending on the need of the simulation algorithm. Nevertheless, not all details of the rooms (scenes 8 – 11) could be captured, and the final models are accompanied by a list of model simplifications and pictures that show the omitted details. Excluded were for example handrails and cable ducts with diameters of a couple of centimeters, stucco ornamentation, projections on walls with a depth of a couple of centimeters, and chandeliers.

5.3.2 Boundary conditions

As detailed in the introduction, it is not feasible to provide high precision boundary conditions in all cases. For this reason, absorption coefficients of the simple scenes 1 – 7 were measured, whereas in case

Material	Absorption coefficient acquisition method (valid freq. range)	Scenes
Medium density fiberboard	ISO 10534-2, normal incidence (100 Hz to 4 kHz)	1 - 7
Stone wool absorber	ISO 10534-2, normal incidence (100 Hz to 4 kHz),	1 - 2
	and angle dependent in situ measurement (300 Hz to 15 kHz)	1 - 2
Wooden diffusor	Angle dependent in situ measurement, (500 Hz to 15 kHz)	1
Tiles of hemi anechoic chamber	Estimated random incidence values (see text)	1, 5 - 7
Room surfaces (22 materials)	Estimated random incidence values (see text)	8-11

Table 5.3: Surface materials of the BRAS database and the applied absorption coefficient acquisition method. The last column lists in which scenes the corresponding materials were used.

of the complex room scenes 8 – 11, values were derived from a combination of measurements and tabulated values of similar materials, which should be considered as approximations. In total, the database contains absorption and scattering coefficients for 37 materials in third octaves from 20 Hz to 20 kHz.

Absorption coefficients

For the scenes covering isolated acoustic phenomena (scenes 1 – 7), three materials were used and described for different configurations. This includes the floor tiles of the hemi anechoic chamber, stone wool absorber tiles, and three medium density fiberboards (MDF) with a thickness of 12 mm (8.91 kg/m^2) and 25 mm (15.53 kg/m^2, and 18.56 kg/m^2), which were used to build the reflecting and diffracting objects in scenes 1 – 7. An overview of all materials and the applied acquisition methods including the valid frequency ranges is shown in Table 5.3.

Normal incidence absorption coefficients were measured according to 10534-2[36] using a circular impedance tube with a diameter of 2 inch. Transfer functions were determined for four microphone positions with distances from the first to the second, third and fourth microphone of 17 mm, 110 mm and 510 mm, respectively. All transfer functions were measured by moving one probe microphone to the desired positions on the mid-axis of the circular tube. Crossover frequencies for the transfer functions between the first (H_{14}) and the second microphone pair (H_{13}), and between the second (H_{13}) and the third microphone pair (H_{12}), were defined at 900 Hz and 1200 Hz, respectively. The measurements are valid above 100 Hz due to the limited frequency range of the loudspeaker, and below 4 kHz due to

[36] ISO 10534-2 (**1998**). *Determination of sound absorption coefficient and impedance in impedance tubes – Part 2: Transfer-function method.*

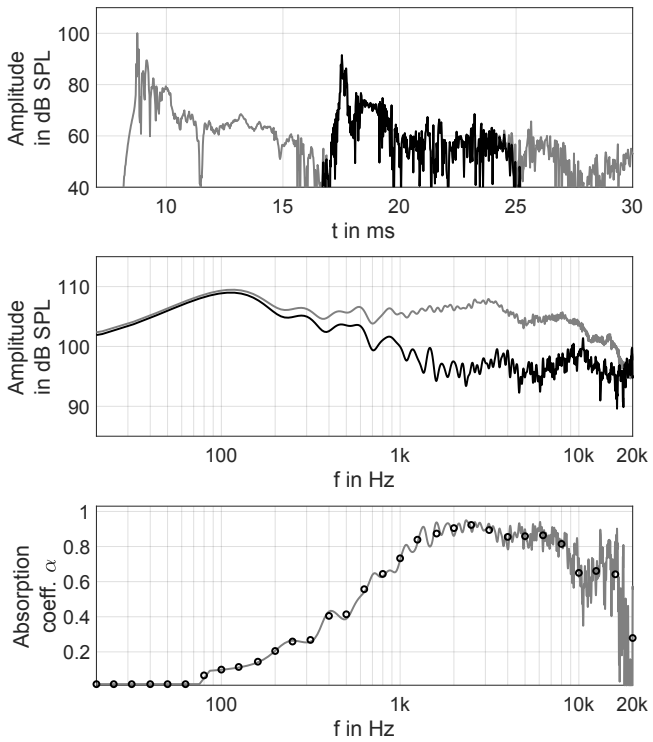


Figure 5.2: Generation of the angle dependent absorption coefficient for the example of the absorbing setup and angles $\phi_{in} = \phi_{out} = 30°$. Top: Windowed (black) and full length impulse response (gray) of the absorbing setup. Middle: Frequency responses of the absorbing (black) and rigid (gray) reflection calculated from the windowed impulse responses. Bottom: Unsmoothed (gray) and final (circles) absorption coefficients α obtained after 3rd octave smoothing.

the diameter of the tube[37].

Angle-dependent absorption coefficients were determined using the setup of scene 1. In the impulse response measurements of the rigid floor, the diffusor and the absorber, the reflection was isolated as described by Mommertz[38]. This was done by a two-sided Hann window (0.6 ms fade in, 1.8 ms fade out, cf. Fig. 5.2, top). The reflection factors were then obtained by spectral division, referencing the windowed impulse responses of the absorber and diffusor to the rigid floor (cf. Fig. 5.2, middle). The absorption coefficients were determined from the absolute value of the reflection factors, smoothed with a one-third octave sliding window (cf. Fig. 5.2, bottom). This was done for the angles $\phi_{in} = \phi_{out} = \{30°, 45°, 60°\}$ based on the impulse responses of the omnidirectional receiver, and for $\phi_{in} = 45°, \phi_{out} = 32°$ based on the binaural data. The lower frequency limit of this method differs for the two measured materials due to edge effects (cf. Table 5.3). Above 15 kHz, minor differences of source, receiver, and probe positions affect the measured result. Missing values below and above the valid frequency ranges were linearly extrapolated or attributed from tabulated values of similar materials[39,40].

Although the reflective floor of the hemi anechoic chamber itself (scenes 1 and 5–7) should be of less relevance for simulation results, the absorption of the tiles were described by an almost frequency independent absorption coefficient with an average value of less than 0.02.

For the complex scenes 8–11, several surfaces were initially measured with a hand-held in situ device[41], consisting of a spherical

[37] ISO 10534-2 (**1998**). *Determination of sound absorption coefficient and impedance in impedance tubes – Part 2: Transfer-function method*.

[38] E. Mommertz (**1995**). "Angle-dependent in-situ measurements of reflection coefficients using a subtraction technique" *Applied Acoustics*.

[39] M. Vorländer (**2008**). *Auralization. Fundamentals of acoustics, modelling, simulation, algorithms and acoustic virtual reality*.

[40] Physikalisch-Technische Bundesanstalt (PTB) (**2012**). `https://www.ptb.de/cms/fileadmin/internet/fachabteilungen/abteilung_1/1.6_schall/1.63/abstab_wf.zip` (last checked Mar. 2019).

[41] M. Müller-Trapet, et al. (**2013**). "On the in situ impedance measurement with pu-probes — Simulation of the measurement setup" *J. Acoust. Soc. Am.*

loudspeaker and a combined sensor unit measuring sound pressure and particle velocity[42]. This method can deliver valid results for normal incidence if applied for porous absorbers in controlled scenarios, but faces several challenges otherwise. Measurements of rather reflective surfaces showed particularly high uncertainties for repeated measurements and different positions. Moreover, measuring ceiling elements, or small but complex objects such as chairs turned out to be impracticable. For these reasons it was chosen to discard the measured absorption values for the surfaces of scenes 8 – 11. Instead, the absorption data was attributed from material databases[43,44] and was derived from in-situ measurements whenever possible (cf. `MaterialOverview.pdf` in the BRAS[45], folder `3 Surface descriptions`). For this purpose, less important materials and small surfaces were disregarded, which lead to four, five, seven, and eight different materials for scenes 8, 9, 10, and 11, respectively. Examples for left out surfaces are small doors far away from all source/receiver positions and stucco ornamentation. These sets of coefficients were termed *initial estimates.*

[42] E. Tijs (**2013**). "Study and development of an in situ acoustic absorption measurement method" Doctoral Thesis.

[43] M. Vorländer (**2008**). *Auralization. Fundamentals of acoustics, modelling, simulation, algorithms and acoustic virtual reality.*

[44] Physikalisch-Technische Bundesanstalt (PTB) (**2012**). `https://www.ptb.de/cms/fileadmin/internet/fachabteilungen/abteilung_1/1.6_schall/1.63/abstab_wf.zip` (last checked Mar. 2019).

[45] L. Aspöck, et al. (**2019**). *BRAS – A Benchmark for Room Acoustical Simulation* `https://dx.doi.org/10.14279/depositonce-6726.2`.

To account for the uncertainty of the initial estimates, a second set of absorption coefficients was provided for the materials of scenes 8 - 11 by fitting the initial estimates to the measured data. For this purpose, the average absorption coefficient $\bar{\alpha}$ was calculated from the initial estimates and measured RIRs

$$\bar{\alpha} = \frac{1}{S} \sum_i S_i \alpha_i \,. \tag{5.1}$$

The surface area $S = \sum_i S_i$ occupied by each material was taken from the 3D room models. In case of the measured RIRs, $\bar{\alpha}$ was calculated under the assumption of a diffuse sound field by solving the Eyring reverberation time for $\bar{\alpha}$

$$T_{\text{Eyring}} = 0.161 \, \frac{V}{-S \; ln(1-\bar{\alpha}) + 4mV} \,, \tag{5.2}$$

(V: room volume according to Table 5.1; m: air attenuation[46]). The second set of coefficients was then obtained by multiplying the initial estimates with the ratio $\bar{\alpha}/\bar{\alpha}'$, where $\bar{\alpha}$ was calculated from the measured RIRs and $\bar{\alpha}'$ from the initial estimates. This procedure was done separately for each third octave frequency and scene (i.e. the same ratio was applied for all materials within a scene). This set of coefficients was termed *fitted estimates*.

[46] H. Kuttruff (**2009**). *Room acoustics*.

Scattering coefficients

As GA based simulation algorithms commonly take into account non-specular reflections, the BRAS also includes random-incidence scattering coefficients for most materials. For this part of the database, no data is based on measurements. Although ISO 17497-1[47] and ISO 17497-2[48] describe standardized measurements of random-incidence scattering and directional diffusion, material samples in the required size could not be removed from the rooms. The scattering data was thus taken from tabulated values of comparable ma-

[47] ISO 17497-1 (**2004**). *Sound-scattering properties of surfaces. Part 1: Measurement of the random-incidence scattering coefficient in a reverberation room.*

[48] ISO 17497-2 (**2012**). *Sound-scattering properties of surfaces. Part 2: Measurement of the directional diffusion coefficient in a free field.*

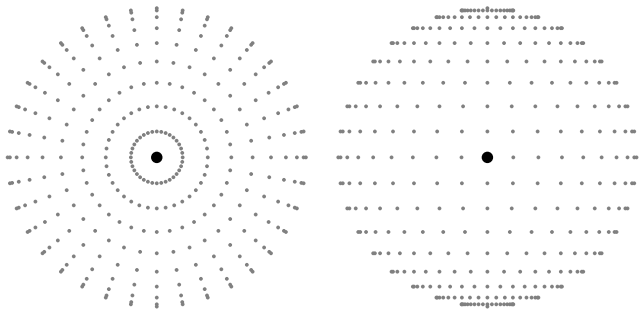


Figure 5.3: $10° \times 10°$ equal angle sampling grids. The frontal direction is marked by the large black dot. Left: Front pole grid that was used for the sources. Right: Top pole grid that was used for the binaural receiver. The resolutions of the actually provided directivities is $1° \times 1°$.

terials (see above) or estimated according to the structural dimensions of the materials[49,50], which are documented in the BRAS by means of scale pictures. Depending on the surface of the material, the scattering coefficients either increase with frequency according to an *S*-shaped curve, or slightly increase for frequencies above 4 kHz in case of the smoother surfaces.

[49] Odeon A/S (**2018**). *ODEON Room Acoustics Software User's Manual (2016. Accessed: Mar. 2019)* `https://odeon.dk/wp-content/uploads/2017/09/ODEON_Manual.pdf`.

[50] J. J. Embrechts and A. Billon (**2011**). "Theoretical determination of the random-incidence scattering coefficients of infinite rigid surfaces with a periodic rectangular roughness profile" *Acta Acust. united Ac.*

5.3.3 *Source Directivities*

The directivities of the Genelec 8020c, QSC K8, and the dodecahedral speaker were measured in the hemi-anechoic chamber at RWTH Aachen University on 2×2 equi angular top pole sampling grid (cf. Fig. 5.3, right) using exponential sweeps (16,384 samples $\approx$ 0.4 s @ 44.1 kHz sampling rate). The sweep length was sufficient to obtain a peak-to-tail SNR of approximately 80 dB (cf. Fig. 5.5). The dodecahedral speaker (cf. Fig. 5.4) is a custom DSP driven 3-way system with a single low-frequency driver operating up to 177 Hz, and mid/high-frequency units consisting of 12 speakers in spherical enclosures (cross-over @ 1.42 kHz). The cross-over frequencies were chosen according to the upper cut-off frequency of the 125 Hz and 1 kHz octave bands. Due to mechanical restrictions, two different systems were used for rotating the transducers: The Genelec 8020c and the dodecahedral speaker were placed on a turntable that controlled the azimuth at a height of 2 m above the ground. The elevation was controlled by an arm that was equipped with a G.R.A.S. 40AF half inch free-field microphone at a distance of 2 m from the speaker. To maintain a sufficient time delay between the direct sound and the floor reflection, the lower hemisphere of the loudspeaker directivities was measured after flipping them about the reference points (cf. Tab. 5.3.1). For both hemispheres, an overlapping region of 4° below the equator was measured for validation. Differences in one-third octave bands between repeated measurements were below 1 dB, except for some directions behind the Genelec 8020c and frequencies above 1 kHz, where less energy is emitted and the measurement is more prone to acoustic noise. In case of the dodecahedral speaker, separate directivities were measured for the mid-frequency and high-frequency unit, in both cases for the full physical setup of the 3-way system (cf. Fig. 5.4), each referenced to the center of the corresponding unit. The low-frequency unit was modeled om-

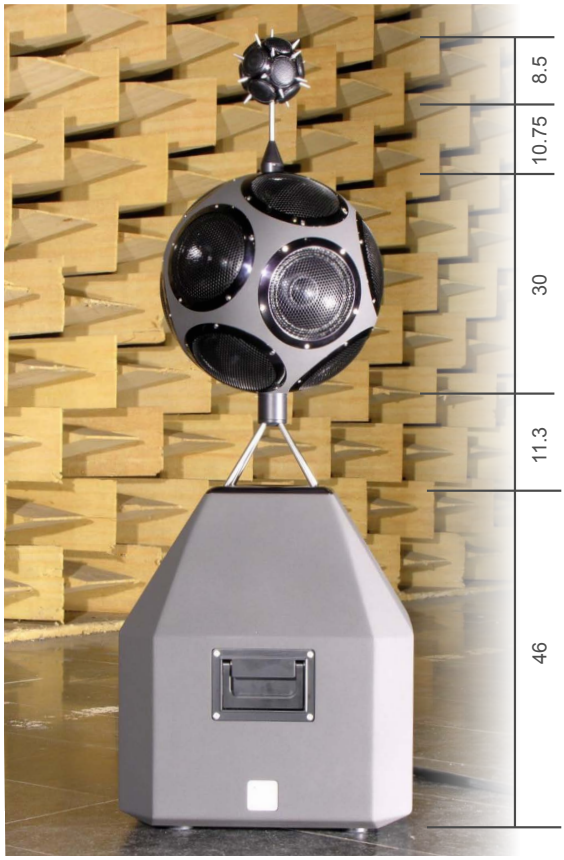


Figure 5.4: Picture of the DSP driven 3-way dodecahedral speaker. Measures are given in cm. A detailed 3D model is contained in the BRAS.

nidirectional because the wavelength at the upper cut-off frequency (1.94 m; 177 Hz) is more than four times larger than the enclosure (0.46 m high). The QSC K8 was rotated using the ELF loudspeaker measurement system (Four Audio) with the microphone placed flush with the floor of the hemi-anechoic chamber at a distance of 8 m. All equalizer settings of the speakers were disabled for reproducibility. Parametric equalizers were used to compensate the free-field on-axis frequency response of the QSC-K8 speaker within a tolerance of $\pm$3.5 dB (cf. Fig. 5.6).

The measurement distances were chosen in agreement with the far field criteria $r \gg l$ and $r \gg fl^2/c$[51], where r is the measurement distance, l the acoustically effective source dimension (tweeter-to-woofer distances of 0.11 m and 0.3 m for the Genelec 8020c and QSC-K8, diameters of 0.3 m and 8.5 cm for the Dodecahedron mid and high frequency units), f the upper frequency limit (16 kHz for the Genelec 8020c, QSC-K8, and Dodecahedron high frequency unit; 1-42 kHz for the Dodecahedron mid frequency unit), and c the speed of sound (343 m/s).

[51] M. Möser (**2009**). *Engineering acoustics: An Introduction to noise control* p. 102.

In post-processing, the common propagation delay was removed and a subsonic high-pass filter was applied (4th order Butterworth -3 dB @ 30 Hz). Due to the different mechanical measurement setups and temporal behaviours, IRs were truncated to 9.5 ms (Genelec), 13.6 ms (dodecahedron), and 80 ms (QSC) to discard reflections by applying a 2 ms Hann window fade in/out (cf. Fig 5.5). In case

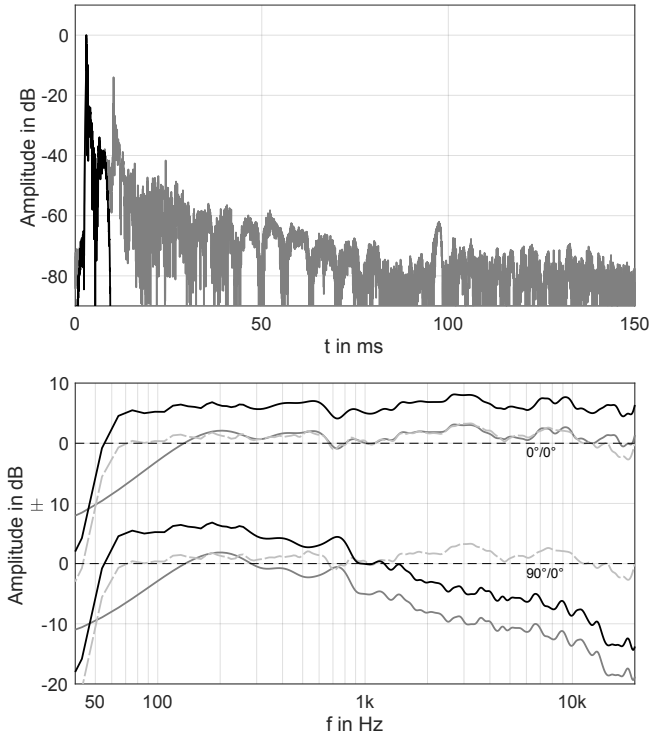


Figure 5.5: Processing of the Genelec 8020c directivity data. Top: Impulse response for the frontal direction (0°/0°) before (gray) and after windowing (black). Bottom: Final frequency responses after spherical harmonics processing (black) in front (0°/0°) and to the side (90°/0°). The measured directivity after windowing (gray) and the single on-axis measurement that was used below 200 Hz (dashed) are given for validation. Data are 12th octave smoothed to improve the visibility.

of the Genelec speaker, the truncation distorted the magnitude response below 200 Hz. To account for this, the magnitude spectrum of a single on-axis measurement – done in the fully anechoic chamber at TU Berlin – was fitted to the truncated IRs by applying a linear fade between 200 Hz and 300 Hz in the frequency domain (gray lines in Fig. 5.5). The phase response did not suffer from the truncation and was thus left unchanged, i.e., was taken from the truncated IRs. The substitution of only the magnitude response is similar to the combination of near-field and far-field loudspeaker measurements[52] As a consequence, the final Genelec directivity is omnidirectional below 200 Hz . Since the originally measured data (before windowing) showed deviations from omnidirectionality of less than 1 dB below 200 Hz, this was considered to be negligible. In a final step, separate spherical harmonics interpolations of the magnitude and unwrapped phase spectra with a spherical harmonics order of 20 were used to increase the spatial resolution (Eq. (3.26) and (3.31) in[53]). Differences before and after the interpolation were smaller than 0.4 dB below 10 kHz, and never exceeded 1 dB. The final data are available as impulse responses and complex one-third octave spectra on a $1^\circ \times 1^\circ$ equal angle front pole sampling grid (cf. Fig. 5.3).

[52] D. B. Keele Jr. (**1974**). "Low-frequency loudspeaker assessment by nearfield sound-pressure measurement" *J. Audio Eng. Soc.*

[53] B. Rafaely (**2015**). *Fundamentals of spherical array processing*.

5.3.4 *Receiver Directivity*

Two types of receivers were used during the acquisition of the BRAS. Impulse responses with omnidirectional receivers were measured either with diffuse field or free field compensated half inch capsules, Brüel&Kjær type 4134 and G.R.A.S. 40AF, respectively, both using Brüel&Kjær 2669-B microphone preamplifiers. They were modeled

as omnidirectional receivers in the BRAS.

The binaural impulse responses were measured with the FABIAN head and torso simulator. FABIAN's acoustically measured head-related impulse responses (HRIRs) were taken from the publicly available FABIAN database[54,55]. It contains HRIRs for 11 different head-above-torso orientations, i.e., head rotations to the left and right in steps of 10° covering the typical range of motion of ±50°[56]. To ensure that head movements can be auralized without any perceivable degradation, the resolution of head rotations was increased to 2° (threshold from[57]). This was done by a perceptually transparent interpolation of HRTFs with identical head-to-source but different torso-to-source angles (termed *original phase, frequency domain, torso interpolation*[58]).

[54] F. Brinkmann, et al. (**2017**c). "A high resolution and full-spherical head-related transfer function database for different head-above-torso orientations" *J. Audio Eng. Soc*

[55] F. Brinkmann, et al. (**2017**b). *The FABIAN head-related transfer function data base* `https://dx.doi.org/10.14279/depositonce-5718.2`.

[56] W. R. Thurlow, et al. (**1967**). "Head movements during sound localization" *J. Acoust. Soc. Am.*

[57] A. Lindau and S. Weinzierl (**2009**). "On the spatial resolution of virtual acoustic environments for head movements on horizontal, vertical and lateral direction" in *EAA Symposium on Auralization*.

[58] F. Brinkmann, et al. (**2015**b). "Audibility and interpolation of head-above-torso orientation in binaural technology" *IEEE J. Sel. Topics Signal Process.*

5.3.5 *Room Impulse Responses*

Room impulse responses were measured using exponential sweeps (262,144 samples ≈ 6 s @ 44.1 kHz sampling rate), and averaged across four repeated measurements to increase the SNR. A high frequency amplitude reduction that can occur when averaging room impulse responses was not observed[59]. Deviations between reverberation times (T_{20} @ 4 kHz and 8 kHz) of averaged and single impulse responses were 0.4% (SD = 0.2%). Scenes 1-7 were measured with a G.R.A.S. 40AF free field compensated microphone capsule; scenes 8–11 with a diffuse field compensated B&K 4134 capsule. The microphones were always oriented towards the current sound source position, corresponding to the 0° direction as defined by the manufacturers. The input measurement chain was calibrated using a B&K Type 4231 sound calibrator, and the input channels of the interfaces were calibrated using a voltage calibrator. The input-to-output latency was removed according to a loopback measurement. To cut the impulse responses to the length defined in the scene description, the last 1024 samples were faded out by application of a one-sided Hann window. Details about the hardware used for each scene and the final impulse response lengths can be found in the database documentation[60].

[59] B. N. J. Postma and B. F. G. Katz (**2016**). "Correction method for averaging slowly time-variant room impulse response measurements" *J. Acoust. Soc. Am.*

For scenes 9–11, source and receiver positions, and characteristics were chosen according to the requirements of ISO-3382-1[61]. The measured IRs using the omnidirectional dodecahedron source thus allow the processing of most of the room acoustical parameters. A Matlab script for calculating C_{80}, D_{50}, EDT or T_{20} (among other parameters) using the *ita_roomacoustics* method of the ITA-Toolbox is part of the BRAS.

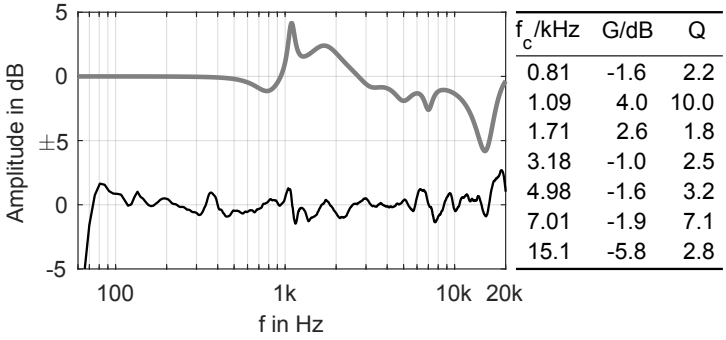
[60] L. Aspöck, et al. (**2019**). *BRAS – A Benchmark for Room Acoustical Simulation* `https://dx.doi.org/10.14279/depositonce-6726.2`.

[61] ISO 3382-1 (**2009**). *Measurement of room acoustic parameters – Part 1: Performance spaces*.

5.3.6 *Binaural Room Impulse Responses*

BRIRs were measured with the FABIAN head and torso simulator[62] using swept sines and spectral deconvolution. The sweep length and level was adjusted to obtain SNRs of about 90 dB for a source in front of FABIAN with the exception of scene 5, where a reduced

[62] A. Lindau, et al. (**2007**). "Binaural resynthesis for comparative studies of acoustical environments" in *122th AES Convention, Convention Paper 7032*.



f_c/kHz	G/dB	Q
0.81	-1.6	2.2
1.09	4.0	10.0
1.71	2.6	1.8
3.18	-1.0	2.5
4.98	-1.6	3.2
7.01	-1.9	7.1
15.1	-5.8	2.8

Figure 5.6: Equalized on-axis magnitude response of the QSC-K8 speaker (black), equalization (gray), and equalizer settings (center frequency, gain, quality). $Q = f_c/|f_2 - f_1|$, where f_1, and f_2 are the two frequencies with a magnitude of G/2 dB (*midpoint dB gain* definition of the bandwidth according to R. Bristow-Johnson (**1994**) "The equivalence of various methods of computing biquad coefficients for audio parametric equalizers." In *97th AES Convention*.

SNR of 70 dB was obtained due to the energy loss caused by the blocked direct sound path. FABIAN is equipped with DPA 4060 miniature electret condenser microphones located at the blocked ear channel entrances, and a computer controllable neck joint (Amtec Robotics PW-070, precision $\pm 0.02°$) that was used to obtain BRIRs for head orientations of $\pm 44°$ in steps of $2°$ to cover the typical range of motion, and allow for an artifact free dynamic auralization of head rotations[63,64].

In scenes 1, 3, 5, and 8, BRIRs were measured for selected position of the Genelec 8020c, while a QSC-K8 speaker was set up at five different positions to mimic a string quartet with a singer in scenes 9–11. For this purpose, the speakers were arranged in a hemi-circle with a radius of 1.5 m at positions of $\pm 90°$ (1st violin, cello) and $\pm 30°$ (2nd violin, viola), and tweeter heights of 103 cm (violins, viola) and 81 cm (cello). To realize the tilting and tweeter heights, the speakers were placed on a chair on top of a box (flight case filled with a layer of porous absorbers). A 3D model of the setup including the box, chair and loudspeaker is contained in the BRAS[65]. The speaker that mimicked the singer was placed on a stand in the center of the semi-circle with a tweeter height of 1.5 m. Except for the center, the speakers were rotated and tilted until the tweeter pointed towards FABIAN's head (the position and orientation of the speaker was identical for all measured head orientations). BRIRs for the five source positions were measured one after another to avoid reflections from other speakers. Parametric equalizers were used to compensate the free-field on-axis frequency response of the QSC-K8 speaker within a tolerance of ± 3.5 dB (cf. Fig. 5.6).

In post-processing, the smallest common propagation delay was removed from all BRIRs of each scene (to minimize the latency in auralizations), a band pass was applied to suppress noise (4th order Butterworth, -3 dB @ 50 Hz/20 kHz), and the on-axis magnitude response of FABIAN's DPA microphones was removed by applying minimum phase inverse filters with a length of 128 samples. For this purpose, the DPA's transfer functions were measured in the fully anechoic chamber of TU Berlin 4.5 m in front of a Genelec 8030a speaker by substitution[66] with a Brüel&Kjær 1/4 inch capsule of Type 4135. Before inverting the magnitude response, it was normalized to 1 (0 dB) by averaging across frequencies between 100

[63] W. R. Thurlow, et al. (**1967**). "Head movements during sound localization" *J. Acoust. Soc. Am.*

[64] A. Lindau and S. Weinzierl (**2009**). "On the spatial resolution of virtual acoustic environments for head movements on horizontal, vertical and lateral direction" in *EAA Symposium on Auralization*.

[65] L. Aspöck, et al. (**2019**). *BRAS – A Benchmark for Room Acoustical Simulation* `https://dx.doi.org/10.14279/depositonce-6726.2`.

[66] IEC 60268-4 (**2014**). *Sound system equipment – Part 4: Microphones.*

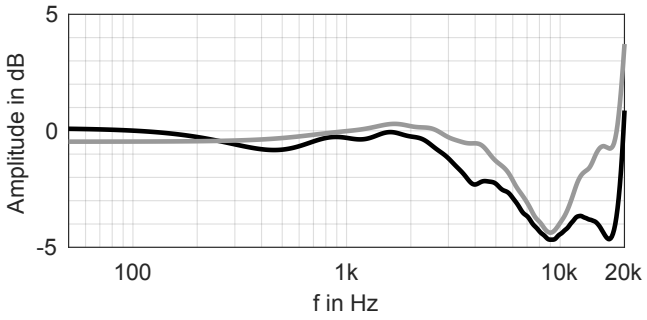


Figure 5.7: Magnitude response of the inverse filter of FABIAN's DPA 4060 microphones (black: left ear, gray: right ear).

Hz and 400 Hz, and linearly extrapolated below 100 Hz to account for the limited frequency range of the Genelec speaker. Finally, the minimum-phase was obtained from the cepstrum[67]. The magnitude response of the final inverse filter is shown in Fig. 5.7. Afterwards the impulse responses were truncated at the position where the noise floor became visible and squared sine fades were applied to avoid discontinuities at the start and beginning. In a last step, all BRIRs of one scene were normalized to the mean absolute level between 300 Hz and 1 kHz taken from the BRIR with neutral head orientation, and the speaker that was closest to FABIAN. To avoid introducing interaural differences the level was averaged across the left and right ear prior to normalization. This way, the relative level differences between BRIRs for different speakers remained. The post-processing was done with Matlab and AKtools[68].

For the audio content provided in the database (anechoic string quartet, see Section 5.4) the bandwidth is sufficient: C2 is the lowest playable note on a Cello and already has a fundamental frequency of 65 Hz. For other stimuli, the 50 Hz limit could indeed be a restriction. However, this restriction would affect the reference as well as the simulation after aplying the same band-pass to the latter.

[67] A. V. Oppenheim and R. W. Schafer (**2010**). *Discrete-time signal processing* , pp. 980.

[68] F. Brinkmann and S. Weinzierl (**2017**). "AKtools – An open software toolbox for signal acquisition, processing, and inspection in acoustics" in *142nd AES Convention, Convention e-Brief 309*.

5.4 Database

The BRAS is available under a Creative Commons share alike license (CC-BY-SA 4.0). For a detailed description of the structure and data format please refer to the documentation[69]. To assure accessibility, the content is provided in open, or wide spread file formats wherever possible: The scene geometries are given in SketchUp files, accompanied by overview and detail photos of the scenes. The source and receiver directivities, as well as the scattering and absorption coefficients (initial and fitted estimates) are provided in text files. The measured IRs are given as wav files and SOFA containers[70]. For convenience, additional data that serves the evaluation of simulated IRs is also provided. This includes text files with room acoustic parameters for scenes 9 – 11 (cf. Section 5.3.5) and an exemplary Matlab script that shows how to calculate the parameters using the ITA-Toolbox[71]. If the provided parameters are used as they are, this script should also be used to calculate parameters from simulated

[69] L. Aspöck, et al. (**2019**). *BRAS – A Benchmark for Room Acoustical Simulation* `https://dx.doi.org/10.14279/depositonce-6726.2`.

[70] AES Standards Comittee (**2015**). *AES69-2015: AES standard for file exchange - Spatial acoustic data file format*.

[71] M. Berzborn, et al. (**2017**). "The ITA-Toolbox: An open source MATLAB toolbox for acoustic measurements and signal processing" in *Fortschritte der Akustik – DAGA 2017*.

impulse responses to avoid uncertainties due to different onset or noise detection procedures[72]. Moreover, a short excerpt of Mozart's string quartet No. 1 (bars 1 – 6 from the second movement) played by the Reinhold Quartet[73] and recorded at the anechoic chamber of TU Berlin is given for the perceptual evaluation of scenes 9 – 11. Compensation filters for common headphone models, and FABIAN's inverted diffuse field transfer function are provided as part of the FABIAN database[74], and can be used for headphone equalization in the context of binaural auralizations. The compensation filters were obtained by inverting the averaged headphone impulse responses (HpIRs) (headphones were re-positioning in between measurements to account for positioning variability). Regulated least mean squares inversion was used to limit the gain of the inverse filter at the frequencies of deep and narrow notches[75].

[72] B. F. G. Katz (**2004**). "International round robin on room acoustical impulse response analysis software" *Acoustic research letters online*.

[73] `www.reinholdquartett.de`, checked Feb. 2018.

[74] F. Brinkmann, et al. (**2017**b). *The FABIAN head-related transfer function data base* `https://dx.doi.org/10.14279/depositonce-5718.2`.

[75] F. Brinkmann, et al. (**2017**c). "A high resolution and full-spherical head-related transfer function database for different head-above-torso orientations" *J. Audio Eng. Soc*

5.5 *Discussion and Outlook*

The presented database of transfer functions for room acoustical scenes ranging from a single reflection on a quasi infinite plate to a large lecture hall is meant to provide a Benchmark for Room Acoustical Simulations (BRAS). With information about the primary structure, the boundary conditions, and the source and receiver characteristics, these transfer functions can be used as a ground truth for acoustical simulation software.

By describing the measurements conducted to create this database, it should be possible to evaluate the related uncertainties and thus the extent up to which differences between simulations and the reference can be reliably attributed to shortcomings in the simulation algorithm itself.

As shown, the source/receiver positions could be established with a precision of about 3.5 cm and the scene geometries were obtained with a precision of about 2.5 cm. The level of detail of the complex scenes 8 – 11 exceeds the resolution of 50 cm recommended for GA simulations[76], and is sufficient for wave based simulations at least up to the Schroeder frequency. The estimation of the surface absorption and scattering properties, however, remains a challenging task. While the applied in situ measurement methods provided reliable results for scenes 1 – 7[77,78,79], uncertainties increase and the valid low-frequency range decreases if they are used in complex environments[80,81] (cf. Section 5.3.2). For scenes 8 – 11, a simplified room representation with partly estimated boundary conditions is thus currently unavoidable, and the BRAS contains random incidence absorption coefficients rather than complex valued reflection factors in these cases. Although neglecting the phase information limits the validity of any simulation to the range above the Schroeder frequency[82], it is unlikely that this causes perceivable artifacts under realistic conditions (c.f. Section III-B in[83]).

The positioning of the sources and receivers is another source of error. As discussed in Section 5.3.1, the uncertainty is uncritical with

[76] M. Vorländer (**2008**). *Auralization. Fundamentals of acoustics, modelling, simulation, algorithms and acoustic virtual reality* , p. 176.

[77] K. Hirosawa, et al. (**2009**). "Comparison of three measurement techniques for the normal absorption coefficient of sound absorbing materials in the free field" *J. Acoust. Soc. Am.*

[78] E. Tijs (**2013**). "Study and development of an in situ acoustic absorption measurement method" Doctoral Thesis.

[79] M. Müller-Trapet, et al. (**2013**). "On the in situ impedance measurement with pu-probes — Simulation of the measurement setup" *J. Acoust. Soc. Am.*

[80] C. Nocke (**2000**). "In-situ acoustic impedance measurement using a free-field transfer function method" *Applied Acoustics*.

[81] E. Brandão, et al. (**2015**). "A review of the in situ impedance and sound absorption measurement techniques" *Acta Acust. united Ac.*

[82] M. Vorländer (**2008**). *Auralization. Fundamentals of acoustics, modelling, simulation, algorithms and acoustic virtual reality.*

[83] C.-H. Jeong, et al. (**2014**). "Influence of impedance phase angle on sound pressures and reverberation times in a rectangular room" *J. Acoust. Soc. Am.*

regard to changes of the overall level and will certainly have a negligible influence on the measured room acoustical parameters[84,85]. However, differences between the geometrical and acoustically estimated loudspeaker positions were up to 5.5 cm, which corresponds to approximately 7 samples at a sampling rate of 44.1 kHz, and a frequency of 6.2 kHz. This has to be kept in mind when comparing temporal and comb filter structures of simulated and measured data.

This initial version of the database, which will be maintained by the related groups at RWTH Aachen and TU Berlin, is meant to serve the community as a resource to be expanded in the future, for example by adding new scenes or by applying advanced measurement methods for acoustical boundary conditions. These extensions, both by the original authors or other groups, can then be published as new versions of the current database. Results of the simulation and auralization round robin campaign conducted between 2016 and 2018 based on the BRAS are published in Chapter 6.

[84] K. Sekiguchi and T. Hanyu (**1998**). "Study on acoustic index variations due to small changes in the observation point" in *15th International Congress on Acoustics*.

[85] I. B. Witew, et al. (**2010**). "Uncertainty of room acoustics measurements - How many measurement positions are necessary to describe the conditions in auditoria?" in *Proc. Int. Symp. on Room Acoustics (ISRA)*.

6
A round robin on room acoustical simulation and auralization

Fabian Brinkmann, Lukas Aspöck, David Ackermann, Steffen Lepa, Michael Vorländer, and Stefan Weinzierl. *J Acust. Soc. Am.* **145**(4), 2746–2760. DOI: 10.1121/1.5096178.
(Accepted manuscript. CC-BY 4.0)

A ROUND ROBIN was conducted to evaluate the state of the art of room acoustic modeling software both in the physical and perceptual realms. The test was based on six acoustic scenes highlighting specific acoustic phenomena and for three complex, "real-world" spatial environments. The results demonstrate that most present simulation algorithms generate obvious model errors once the assumptions of geometrical acoustics are no longer met. As a consequence, they are neither able to provide a reliable pattern of early reflections nor do they provide a reliable prediction of room acoustic parameters outside a medium frequency range. In the perceptual domain, the algorithms under test could generate mostly plausible but not authentic auralizations, i.e., the difference between simulated and measured impulse responses of the same scene was always clearly audible. Most relevant for this perceptual difference are deviations in tone color and source position between measurement and simulation, which to a large extent can be traced back to the simplified use of random incidence absorption and scattering coefficients and shortcomings in the simulation of early reflections due to the missing or insufficient modeling of diffraction.

6.1 Introduction

Room acoustical simulation shows an increasing number of applications covering not only the classical tasks of acoustical and electroacoustical planning[1], but also fields such as architectural history[2,3], music research[4], game audio[5], or virtual acoustic reality in general[6]. This applies to wave-based simulations as well as to simulations based on geometrical acoustics (GA)[7] or hybrid approaches[8]. Many of these applications make use of the possibility to generate binaural signals based on including head-related transfer functions (HRTFs)

[1] H. Kuttruff (**1991**). "Digital simulation of concert hall acoustics and its applications" *Acoustic Bulletin*.

[2] J. H. Rindel (**2011**b). "The ERATO project and its contribution to our understanding of the acoustics of ancient theatres" in *The Acoustics of Ancient Theatres Conference*.

[3] S. Weinzierl, et al. (**2015**). "The acoustics of Renaissance theatres in Italy" *Acta Acust. united Ac.*

[4] Z. S. Kalkandjiev and S. Weinzierl (**2015**). "The influence of room acoustics on solo music performances: An empirical investigation" *Psychomusicology: Music, Mind, and Brain*.

[5] C. Schissler, et al. (**2016**). "Efficient hrtf-based spatial audio for area and volumetric sources" *IEEE Transactions on Visualization and Computer Graphics*.

[6] M. Vorländer (**2008**). *Auralization. Fundamentals of acoustics, modelling, simulation, algorithms and acoustic virtual reality*.

[7] L. Savioja and U. P. Svensson (**2015**). "Overview of geometrical room acoustic modeling techniques" *J. Acoust. Soc. Am.*

[8] A. Southern, et al. (**2013**). "Room impulse response synthesis and validation using a hybrid acoustic model" *IEEE Transactions on Audio, Speech, and Language Processing*.

into the numerical signal chain, a process which was coined auralization[9]. At the same time, there is no undivided confidence in the reliability of room acoustical simulations when it comes, for example, to the design of new performance venues for music and speech, where acoustic scale models are still an important tool with specific advantages[10]. The application and further development of room acoustic simulations is thus crucially dependent on the availability of a procedure to objectively assess the accuracy of these applications – the more so since background theories such as GA make obvious simplifications which are valid only in a limited frequency range.

There have been different attempts to validate the mentioned modeling approaches and the related software implementations. Two databases with analytically defined test scenarios were established by Otsuru *et al.*[11] and Hornikx *et al.*[12]. They are intended for cross-validation of wave based simulation algorithms and are not dependent on measured reference data. This is an approach that guarantees a perfect reference, but a viable option only for very simple scenes for which analytic solutions are available.

In three round robin experiments conducted between 1994 and 2002 (RR-I to RR-III)[13,14,15,16] the results of different room acoustical simulation algorithms were compared to measurements of a smaller lecture hall (RR-I), a multipurpose hall (RR-II), and a music studio (RR-III). In this series of tests, different information was provided to the participants at different phases. In phase I of RR-I and RR-II, the participants had to estimate the geometry and the boundary conditions themselves from architectural plans and written information ("3 mm carpet"); in phase II the data were harmonized based on a common three-dimensional (3D) model and boundary conditions estimated by room acoustical measurements. In RR-III, absorption and scattering coefficients for one wall and the ceiling of the room were measured in the reverberation room, and taken from tabulated data otherwise. Measured and simulated room impulse responses (IRs) were compared based on room acoustical parameters, i.e., audio features extracted from energy decay representations, such as the early decay time (EDT) and other parameters suggested in ISO 3382-1[17].

The biggest challenge in working with measured references in these tests has been to guarantee an exact match of the measured situation and input parameters of the numerical model. This applies to the geometric model of the acoustic scene, the behaviour of the sources and receivers as an integral part of the acoustic transfer path, and – above all – to the acoustic boundary conditions. For complex rooms such as concert venues or lecture halls, a comprehensive specification of absorption and scattering for all boundaries is practically impossible because neither can all different surfaces with their different types of installation be measured in the laboratory nor are any (standardized) full-range measurement techniques available to determine them in situ[18]. Fitting the input parameters according to measurements of the reverberation time, on the other hand, may be a pragmatic and often applied solution in room acous-

[9] M. Kleiner, et al. (**1993**). "Auralization – An overview" *J. Audio Eng. Soc.*

[10] J. H. Rindel (**2011**a). "Room acoustic modelling techniques: A comparison of a scale model and a computer model for a new opera theatre" *Building Acoustics*.

[11] T. Otsuru, et al. (**2005**). "Constructing a database of computational methods for environmental acoustics" *Acoust. Sci. & Tech.*

[12] M. Hornikx, et al. (**2015**). "A platform for benchmark cases in computational acoustics" *Acta Acust. united Ac.*

[13] M. Vorländer (**1995**). "International round robin on room acoustical computer simulations" in *15th International Congress on Acoustics*.

[14] I. Bork (**2000**). "A comparison of room simulation software – The 2nd round robin on room acoustical computer simulation" *Acta Acust. united Ac.*

[15] I. Bork (**2005**a). "Report on the 3rd round robin on room acoustical computer simulation - Part I: Measurements" *Acta Acust. united Ac.*

[16] I. Bork (**2005**b). "Report on the 3rd round robin on room acoustical computer simulation - Part II: Calculations" *Acta Acust. united Ac.*

[17] ISO 3382-1 (**2009**). *Measurement of room acoustic parameters – Part 1: Performance spaces*.

[18] M. Vorländer (**2013**). "Computer simulations in room acoustics: Concepts and uncertainties" *J. Acoust. Soc. Am.*

tic planning. As a procedure for the evaluation of numerical simulations, however, it would contain an element of circular reasoning if both the premises (the boundary conditions) and the success of the simulation were determined by the measurement of the same room acoustical parameters, or by ones that strongly correlate with each other. Hence, although RR-I to RR-III imitated a "real-world" acoustical planning scenario and gave an impression of how reliable different room acoustics simulation softwares are as planning tools, they could hardly give concrete insights into the strengths and weaknesses of the algorithms themselves.

A reliable reference for room acoustical simulations with respect to geometry and boundary conditions can only be provided if the scene is sufficiently simple so that the relevant measuring methods do not reach their limits. This approach was followed by Tsingos *et al.* when setting up the "Bell Labs Box", i.e., a 16 m^3 rectangular enclosure with one baffle inside, in order to compare measured and simulated IRs and to validate a proprietary simulation algorithm[19]. The planned extension of the test system toward different and more complex configurations and an evaluation of different numerical simulations, however, has not yet taken place.

[19] N. Tsingos, et al. (**2002**). "Validating acoustical simulations in the Bell Labs Box" *IEEE Computer Graphics and Applications*.

The round robin on room acoustical simulation and auralization (RR-SA) presented here represents a combination of both approaches and extends them to an evaluation in the physical *and* perceptual realm. The test was based on a database of measured IRs established for this purpose[20,21]. It contains 3D room models, source and receiver directivities, and one-third octave absorption and scattering coefficients for 11 acoustic scenes (Table 6.2.1). Eight of these scenes are simple configurations for which all parameters could be measured in the laboratory with high precision. They were designed to isolate specific acoustical phenomena such as single and multiple reflections on finite and infinite plates, scattering, diffraction, the seat dip effect, or a coupled room. Three of the scenes are complex, real-world rooms similar those used in RR-I to RR-III for which only a best possible, practical estimate of the parameters could be given. A selection of these scene descriptions was provided to developers of room acoustic simulation software who were given six month to simulate IRs based on the provided data.

[20] L. Aspöck, et al. (**2018**). *GRAS – Ground Truth for Room Acoustical Simulation* `https://dx.doi.org/10.14279/depositonce-6726`.

[21] Please note that an updated version named BRAS has been published in the meantime: L. Aspöck, et al. (**2019**). *BRAS – A Benchmark for Room Acoustical Simulation* `https://dx.doi.org/10.14279/depositonce-6726.2`.

To evaluate the results of the numerical simulations in the physical domain, measured and simulated IRs were compared based on temporal and spectral features. Moreover, dynamic auralizations of the simulated scenes based on binaural room impulse responses (BRIRs) were evaluated against their measured counterparts with the simulation using HRTFs corresponding to the binaural receiver which was also used for the measurements. The listening test yielded measures for the plausibility and authenticity of the simulation, as well as difference ratings for a selection of specific perceptual qualities.

#	Scene	Algorithms
1	Single reflection (infinite plate)	5/-
2	Single reflection (finite plate)	4/-
3	Multiple reflections (parallel finite plates)	5/-
4	Single reflection (reflector array)	3/-
5	Diffraction (infinite wedge)	3/-
6	Diffraction (finite body)	-/-
7	Multiple diffraction (seat dip effect)	-/-
8	Coupled rooms	6/-
9	Small room (seminar room)	6/4
10	Medium room (chamber music hall)	6/4
11	Large room (auditorium)	6/4

Table 6.1: Overview of the 11 scenes contained in the database: Scenes 1 – 8 are designed scenarios to isolate acoustical phenomena and scenes 9 – 11 are representative room acoustic scenarios. Most scenes include multiple source/receiver positions and configurations (e.g. different surfaces materials). The column *Algorithms* shows the number of participating teams in the physical/perceptual evaluations of the round robin. Gray entries were not considered in the round robin.

6.2 Method

6.2.1 Scene descriptions

For the round robin, 9 of the 11 scenes of the database were selected (Table 6.2.1), each of which was supposed to be simulated with different settings (boundary conditions, source and receiver positions). A short overview of these configurations will be given below, while a more comprehensive description, including all scene configurations, is available in the supplemental materials[22] and in documentation of the database itself[23,24].

Scene 1 realizes a single reflection on quasi infinite rigid, absorbing, and diffusing baffles for incident and exit angles of 30°, 45°, and 60°. **Scene 2** is a single reflection on a finite quadratic plate with edge lengths of 1 m and 2 m, and incident/exit angles of 30°, 45°, and 60°. Receiver positions behind the plate are included to assess diffraction around the reflector. **Scene 3** constitutes a flutter echo between two finite reflectors with edge lengths of 2 m and a single source-receiver configuration. **Scene 4** realizes a single reflection on an array of nine reflectors with edge lengths of 68 cm (spaced 13 cm apart) for incident and exit angles of 30°, 45°, and 60°, as well as a reflection point on the center of a reflector and a reflection point between four reflectors. **Scene 5** features the diffraction around a quasi infinte wedge (4.75 m × 2.07 m) for four different source and receiver heights below and above the upper edge of the wedge. **Scene 8** establishes the double sloped energy decay of a reverberation chamber coupled to a laboratory room. Different degrees of coupling were realized by two opening angles of the connecting door (4.1° and 30.4°) and source positions inside both rooms. **Scenes 9–11** are complex real-life environments of different size where omnidirectional source and receiver configurations according to ISO 3382-1[25] were included, as well as a binaural receiver and directional sources.

In all scenes, IRs were measured with a Genelec 8020c studio monitor and 1/2 inch pressure microphones (G.R.A.S. 40AF, Bruel&Kjær Type 4134). BRIRs were measured with QSC-K8 PA speakers and the FABIAN head and torso simulator[26] (HATS). The QSC-K8 speakers

[22] See Supplementary materials at `http://dx.doi.org/10.14279/depositonce-8510`.

[23] L. Aspöck, et al. (**2018**). *GRAS – Ground Truth for Room Acoustical Simulation* `https://dx.doi.org/10.14279/depositonce-6726`.

[24] Please note that an updated version named BRAS has been published in the meantime: L. Aspöck, et al. (**2019**). *BRAS – A Benchmark for Room Acoustical Simulation* `https://dx.doi.org/10.14279/depositonce-6726.2`.

[25] ISO 3382-1 (**2009**). *Measurement of room acoustic parameters – Part 1: Performance spaces*.

[26] A. Lindau, et al. (**2007**). "Binaural resynthesis for comparative studies of acoustical environments" in *122th AES Convention, Convention Paper 7032*.

were chosen due to their higher sound power enabling auralizations with a higher signal-to-noise ratio; the FABIAN HATS was chosen due to the ability to automatically measure BRIRs for different head orientations. For the complex rooms (scenes 9–11), IRs were additionally measured with an ISO-3382-1[27] compliant dodecahedron speaker for the ISO-compliant analysis of room acoustical parameters.

Five sources were used for the binaural measurements of which four were arranged in a semicircular setup to mimic the positions of a string quartet, and the fifths source was placed in the center of the virtual quartet to mimic the position of a singer. The receiver was placed at a distance of $3-4$ times the critical distance to emphasize the influence of the room (cf. Table 6.2.3). BRIRs were measured for head-above-torso orientations to the left and right in the range of $\pm 44^\circ$ with a resolution of 2°, allowing for a perceptually transparent switching of BRIRs for different head orientations[28] within the typical range of motion[29].

[27] ISO 3382-1 (**2009**). *Measurement of room acoustic parameters – Part 1: Performance spaces*.

[28] A. Lindau and S. Weinzierl (**2009**). "On the spatial resolution of virtual acoustic environments for head movements on horizontal, vertical and lateral direction" in *EAA Symposium on Auralization*.

[29] W. R. Thurlow, et al. (**1967**). "Head movements during sound localization" *J. Acoust. Soc. Am.*

6.2.2 Simulation algorithms

Six teams using five different simulation algorithms participated in the round robin:

BRASS (Brazilian Room Acoustic Simulation Software) is a ray tracing algorithm developed in the academic environment, which clusters reflections up to fifth order to provide accurate early reflections without deploying an image source model[30]. **EASE** V4.4 is a commercial tool for the simulation of room acoustical and electroacoustical environments, which uses image sources for the direct sound and early reflections and ray tracing for the late reverberation[31]. **ODEON** Combined is a commercial tool for room acoustical simulation based on a hybrid ray tracing approach for detecting early specular reflections and calculating late reverberation[32]. **RAVEN** (Room Acoustics for Virtual Environments) is a hybrid algorithm developed in the academic environment that uses image sources for the direct sound and early reflections, as well as ray tracing for the late reverberation[33]. **RAZR** is an open source academic algorithm for the simulation of rectangular rooms through a combination of image sources and a feedback delay network for late reverberation[34].

All algorithms consider frequency dependent absorption and scattering coefficients, air absorption, and arbitrary receiver and source directivities – with the exception of RAZR, which assumes omnidirectional sources, and does not account for scattering. The simulation of diffraction is only implemented in ODEON, and is only activated in case of a blocked direct sound path by estimating diffraction paths around objects. ODEON is also the only algorithm that considers the energy loss of specular reflections caused by diffraction around finite objects, by adjusting the scattering coefficient depending on the incident angle and size of the reflecting surface (cf. manual, pp. 79 and 83[35]). Moreover, ODEON takes into account angle dependent

[30] J. C. B. Torres, et al. (**2018**). "Comparative study of two geometrical acoustic simulation models" *J. Brazilian Soc. of Mechanical Sciences and Engineering*.

[31] Ahnert Feistel Media Group. *EASE – User's Guide & Tutorial (Version 4.3), Acesssed in November 2018* `http://www.afmg-support.de/SoftwareDownloadBase/AFMG/EASE/EASE_4.3_Tutorial_English.pdf` (Accessed: Mar. 2019).

[32] Odeon A/S (**2018**). *ODEON Room Acoustics Software User's Manual (2016. Accessed: Mar. 2019)* `https://odeon.dk/wp-content/uploads/2017/09/ODEON_Manual.pdf`.

[33] D. Schröder (**2011**). "Physically based real-time auralization of interactive virtual environments" Doctoral Thesis.

[34] T. Wendt, et al. (**2014**). "A computationally-efficient and perceptually-plausible algorithm for binaural room impulse response simulation" *J. Audio Eng. Soc*

[35] Odeon A/S (**2018**). *ODEON Room Acoustics Software User's Manual (2016. Accessed: Mar. 2019)* `https://odeon.dk/wp-content/uploads/2017/09/ODEON_Manual.pdf`.

	V/m^3	T_m/s	f_s/Hz	d_c/m	d/m
small	145	2.0	234	0.49	4.00
medium	2,350	1.3	47	2.42	9.95
large	8,650	2.1	31	3.66	11.33

Table 6.2: Selected properties of the small, medium, and large room: Approximate volume V and reverberation time T_m (averaged across 500 Hz and 1 kHz octaves), as well as the corresponding Schroeder frequency $f_s = 2000\sqrt{T/V}$ and critical distance $d_c \approx 0.057\sqrt{V/T}$ for each room. The distance from the binaural receiver to the center QSC-K8 speaker is given by d.

absorption by modifying random incidence coefficients based on the mid-range absorption between 1 and 4 kHz and idealized absorber models (cf. manual, pp. 74[36]).

All simulations were carried out in the groups or companies of the software developers themselves. A second contribution using ODEON 12 came from the Department of Industrial Engineering, University of Bologna. Please note that RAVEN is developed at RWTH Aachen University, which was also involved in acquiring the acoustic scenes that were used in the round robin. However, neither did RAVEN play a role in the generation of the reference database nor were results from RAVEN adjusted according to measured data. To avoid a bias, the RAVEN simulations were conducted by a person who was not aware of the measurement results nor was he involved in the round robin otherwise.

More teams showed interest in contributing to the round robin. Developers of wave-based algorithms, however, were not ready to provide results for the entire audible bandwidth, and commercial algorithms sometimes missed an interface for including other than the stock HRTFs in their simulations. RAZR was allowed to participate despite the high degree of simplification of the underlying algorithm due to the open nature of the call that initiated the round robin. In retrospect, the results turned out to be particularly interesting because they were in some properties well comparable to results from the remaining algorithms.

In the following, the terms *algorithm* and *software* will refer to the combination of the actual simulation software and the people that used it to simulate the IRs.

6.2.3 *Task and data processing*

The participants were instructed to simulate IRs without changing the source and receiver directivities or the surface properties (absorption, scattering). For the simple scenes used (scenes 1–5), the boundary conditions could be reliably determined by laboratory measurements, so no modification would be reasonable. For the complex scenes (scenes 8–11), the measured absorption and scattering coefficients can only be considered as best possible estimations, so the simulation could probably have been improved by fitting the boundary conditions according to the measured results of room acoustical parameters. In this case, the task of the round robin corresponds to the predictive situation of a new room acoustic design where no such measurements are available. To ensure this, the measured IRs were not available to the participants at the time of the simulations. In contrast, the room geometry provided by the 3D model could be

[36] Odeon A/S (**2018**). *ODEON Room Acoustics Software User's Manual (2016. Accessed: Mar. 2019)* `https://odeon.dk/wp-content/uploads/2017/09/ODEON_Manual.pdf`

simplified if required by the specific simulation algorithm since the authors consider such a pre-processing, which is always necessary for high resolution architectural models, as part of the simulation itself.

The source directivities of the Genelec 8020c and QSC-K8 were provided by means of IRs and third octave spectra on a 1×1 equal angle sampling grid. Head-related impulse responses (HRIRs) of the head and torso simulator that was also used for measuring BRIRs were obtained on a 1×1 equal angle sampling grid from the FABIAN database[37,38]. The frequency response of the sources is contained in the measured IRs and BRIRs and the corresponding directivities, while the frequency responses of FABIAN's DPA 4060 microphones are removed from the HRIRs provided and from all measured BRIRs.

The software teams reported that they used the provided directivities without changes, with the exception of RAZR that used omnidirectional sources with only the on-axis frequency response taken from the provided data. The ODEON contribution of the developers' group used a spatial resolution of 10° for the source directivities and 3° for the binaural receiver. Both ODEON contributions converted the directivity information to octave values and restricted the range to center frequencies between 64 Hz and 8 kHz. The contributions from ODEON and RAZR also converted the absorption coefficients to octave values in the ranges from 63 Hz to 8 kHz, and 250 Hz to 4 kHz, respectively. While RAZR neglected the provided scattering coefficients, the ODEON teams obtained a nominal mid-frequency scattering coefficient by averaging between 400 Hz and 1.25 kHz (developers) and by using the provided values at 800 Hz (University of Bologna). None of the teams reported to have changed the 3D models, with the exception of RAZR, which generated rectangular rooms with equivalent volumes maintaining the ratios of the main room dimensions. Most of the teams used a transition order of two and three to combine early and late reflections, with the exception of BRASS, which clustered ray traced reflections up to order five to obtain image-source-like components, and the ODEON contribution from the University of Bologna that used a transition order of ten.

[37] F. Brinkmann, et al. (**2017**c). "A high resolution and full-spherical head-related transfer function database for different head-above-torso orientations" *J. Audio Eng. Soc*

[38] F. Brinkmann, et al. (**2017**b). *The FABIAN head-related transfer function data base* `https://dx.doi.org/10.14279/depositonce-5718.2`.

6.2.4 *Physical evaluation*

For the physical evaluation, IRs for the omnidirectional receiver were processed in MATLAB using methods from the open source project ITA-Toolbox[39]. The measured and simulated IRs were temporally aligned, normalized to the root mean square of the IR, and truncated to a length of 46 ms using a two-sided Hann window (3 ms fade in, 10 ms fade out). In case of scene 8, IRs were truncated to 2.2 s with a 50 ms fade out, and the energy decay curve (EDC) was calculated for the 1 kHz octave band.

For scenes 9–11, room acoustical parameters were calculated according to ISO 3382-1[40] based on IRs measured with the dodecahedral loudspeaker and omnidirectional microphone for two source

[39] M. Berzborn, et al. (**2017**). "The ITA-Toolbox: An open source MATLAB toolbox for acoustic measurements and signal processing" in *Fortschritte der Akustik – DAGA 2017*.

[40] ISO 3382-1 (**2009**). *Measurement of room acoustic parameters – Part 1: Performance spaces*.

and five receiver positions. The parameters for the measured and simulated IRs were calculated using the `ita_roomacoustics` routine. For the simulated RIRs, the calculation of the EDC was based on the entire RIR, while the measured RIRs were truncated after detecting the noise floor according to ISO 3382-1[41]. For the sake of brevity, only T_{20} results are presented as averages over all source/receiver combinations.

Asking the participants for RIRs instead of room acoustic parameters increases the flexibility for further parameter evaluation and ensures identical processing of the RIRs, as different evaluation methods can lead to variations, especially for lower frequency bands[42,43].

Measured and simulated BRIRs were further analyzed with respect to differences in perceived tone color. This was assessed by means of energetic differences in 37 auditory filter bands between 80 Hz and 16 kHz using the gammatone filterbank from the auditory toolbox[44].

[41] M. Guski and M. Vorländer (**2014**). "Comparison of noise compensation methods for room acoustic impulse response evaluations" *Acta Acustica united with Acustica*.

[42] B. F. G. Katz (**2004**). "International round robin on room acoustical impulse response analysis software" *Acoustic research letters online*.

[43] D. Cabrera, et al. (**2016**). "Calculating reverberation time from impulse responses: A comparison of software implementations" *Acoustics Australia*.

[44] M. Slaney (**1998**). "Auditory toolbox. version 2" Technical Report #1998-010 .

6.2.5 *Auralization*

Dynamic auralizations, considering head movements of the listeners in a horizontal range of $\pm 44^\circ$, were obtained by dynamic convolution of the measured and simulated BRIRs with anechoic audio content: The BRIRs for all head orientations and sources were stored in SOFA files[45] and loaded by a customized version of the Sound Scape Renderer[46] (SSR) used for convolution. The BRIRs were selected according to the current head orientation of the listener as provided by a Polhemus Patriot head tracker (precision 0.003°). Dynamic auralizations were used to improve the degree of realism, as it was shown that head movements improve externalization[47] and are naturally used when judging acoustic qualities[48]. Pure Data[49] was used to start and stop the anechoic audio content according to open sound control messages triggered via MATLAB based user interfaces. Pure Data and the SSR ran on a Linux-based desktop computer where the audio routing was done by the Jack Audio Connection Kit (JACK). The user interfaces ran on a separate laptop computer with Windows. The setup made it possible to switch between auralizations rendered from measured BRIRs, and BRIRs from different acoustic simulation algorithms at any time, whereby the audio content was restarted. For playback, Sennheiser HD 800 headphones were used at a playback level of 70 dB(A) (measured with pink noise). To minimize the influence of the headphone, a compensation filter was designed using regularized inversion[50].

Because the BRIRs differed in level across algorithms and compared to the measured data, they had to be normalized. The gain for normalization was obtained by averaging the logarithmic magnitude response of the binaural transfer functions (center source and neutral head orientation of FABIAN) between 200 Hz and 1 kHz, and across the left and right ear. One gain value was applied to all BRIRs of each algorithm, assuming that the algorithms preserved the

[45] AES Standards Comittee (**2015**). *AES69-2015: AES standard for file exchange - Spatial acoustic data file format*.

[46] M. Geier, et al. (**2008**). "The sound scape renderer: A unified spatial audio reproduction framework for arbitrary rendering methods" in *124th AES Convention, Preprint 7330*.

[47] E. Hendrickx, et al. (**2017**). "Influence of head tracking on the externalization of speech stimuli for non-individualized binaural synthesis" *J. Acoust. Soc. Am.*

[48] C. Kim, et al. (**2013**). "Head movements made by listeners in experimental and real-life listening activities" *J. Audio Eng. Soc.*

[49] M. S. Puckette et al. (**1997**). "Pure data." in *Int. Computer Music Conf. (ICMC)*.

[50] A. Lindau and F. Brinkmann (**2012**). "Perceptual evaluation of headphone compensation in binaural synthesis based on non-individual recordings" *J. Audio Eng. Soc.*

level difference between the sources and ears, which was confirmed by an analysis of the level across source positions. Afterwards, the authors made manual adjustments in the range of ± 0.5 dB to optimize the loudness matching between algorithms and across measured and simulated data by means of informal listening (cf. audio examples[51]).

[51] See Supplementary materials at `http://dx.doi.org/10.14279/depositonce-8510`.

6.2.6 *Perceptual evaluation*

The perceptual evaluation was done based on two measures for the overall perceived difference between measurement and simulation (*authenticity*, *plausibility*), and a differential diagnosis using the *Spatial Audio Quality Inventory* (SAQI), a qualitative test including 48 perceptual qualities relevant for the quality of virtual acoustic environments[52].

A measure for *authenticity*, indicating the existence of any audible difference between measurement and simulation[53], was obtained by implementing a two interval, two alternative forced choice test (2I/2AFC) as a double-blind and criterion-free procedure. On a user interface with three buttons A, B, and X, the subjects were asked "Does X equal A of B?". Reference and simulation were randomly assigned to the buttons, and the participants could listen to A, B, and X in any order and as often as they wanted before making their choice.

[52] A. Lindau, et al. (**2014**b). "A Spatial Audio Quality Inventory (SAQI)" *Acta Acust. united Ac.*

[53] F. Brinkmann, et al. (**2017**a). "On the authenticity of individual dynamic binaural synthesis" *J. Acoust. Soc. Am.*

To analyze the significance of the results, the type I error level (concluding that there is a difference although there is none) and the type II error level (concluding that there is no difference although there is one) were both set to 0.05. Since testing for authenticity requires proving the null hypothesis (no audible difference), which is not possible with inferential statistics, a minimum-effect test was conducted based on a practically meaningful detection rate of 0.9[54]. Hence, the alternative hypothesis to be rejected for assuming authenticity was $H_1(p_{\text{2AFC}} \geq 0.9)$, and the null hypothesis (no difference) was $H_0(p_{\text{2AFC}} = 0.5)$ with 0.5 as the 2AFC guessing probability. According to the desired error levels and effect size, $N = 13$ trials had to be conducted per participant[55], with authenticity to be assumed in the case of less than $N_{\text{min}} = 10$ correct answers. Note that $N_{\text{crit.}} = 10$ refers to a detection rate of about 75%, which equals a guessing probability of 50% and is the definition of the just noticeable difference (JND).

[54] K. R. Murphy and B. Myors (**1999**). "Testing the hypothesis that treatments have negligible effects: Minimum-effect tests in the general linear model" *J. of Applied Psychology*.

[55] L. Leventhal (**1986**). "Type 1 and type 2 errors in the statistical analysis of listening tests" *J. Audio Eng. Soc.*

A looped pink noise pulse between 100 Hz and 20 kHz and a duration of 1 s (20 ms squared sine ramps) followed by 1.5 s silence was used as audio content due to its high potential to reveal possible flaws of the simulations that are related to timbral and spatial perceptions. The bandwidth was chosen according to the operating range of the measurement equipment and the frequency range where absorption and scattering coefficients were provided. The pulse was auralized by the rightmost source as viewed from the direction of the binaural receiver (position of the cello in the virtual string quartet).

As a somewhat less strict criterion *plausibility* was determined, indicating whether BRIRs can be identified as "simulated" according to artefacts in the stimulus itself, i.e., without immediate comparison to an external reference. The test was implemented as a yes–no task. After each presentation, participants were asked "Was this an audio example from a real room?", and the answers were analyzed with signal detection theory (SDT)[56]. This allows to obtain a criterion-free measure for the sensory difference d' between auralizations based on measured and simulated BRIRs, with $d' = 0$ indicating that differences were inaudible and $d' > 0$ indicating that differences are audible. The sensory difference can be converted to the easier to interpret 2AFC detection rate by $p_{2\text{AFC}} = \Phi(d'/\sqrt{2})$, where $\Phi(\cdot)$ is the cumulative standard normal distribution.

[56] A. Lindau and S. Weinzierl (**2012**). "Assessing the plausibility of virtual acoustic environments" *Acta Acust. united Ac.*

In analogy to authenticity, plausibility was tested separately for each participant. To analyze the significance of the results, the type I error level (wrongly concluding that a simulation is *not* plausible) and the type II error level (wrongly concluding that a simulation is plausible) were again balanced and set to 0.05. According to the desired error levels, the meaningful d'_{min} to be rejected in a minimum-effect test[57], is $d'_{\text{min}} = 0.82$ (c.f. Eq. (13) in Lindau and Weinzierl[58]). It corresponds to a 2AFC detection rate of $p_{2\text{AFC}} = 0.72$, which is similar to the critical value of the test for authenticity.

[57] K. R. Murphy and B. Myors (**1999**). "Testing the hypothesis that treatments have negligible effects: Minimum-effect tests in the general linear model" *J. of Applied Psychology*.

[58] A. Lindau and S. Weinzierl (**2012**). "Assessing the plausibility of virtual acoustic environments" *Acta Acust. united Ac.*

For the test, auralizations of 3 – 5 seconds duration were presented to the participants. The presentation order was randomized, and participants did not know whether an auralization was based on measured or simulated BRIRs, but were informed that the test conditions were approximately evenly distributed across $N = 100$ test trials (5 source positions $\times$ 20 audio contents). To avoid possible familiarization, 20 different monophonic audio contents were used exactly once with each of the 5 sources. These included an artificial noise signal, female/male speech and singing in different languages, solo instrument recordings, and excerpts of different pop songs. A visual impression of the room was provided by a 55" curved screen with a two picture slide show. One picture showed the entire room with an empty stage, and one was taken from the virtual listening position with loudspeakers on the stage (cf. SuppPub3[59], Fig. S3-34).

[59] See Supplementary materials at `http://dx.doi.org/10.14279/depositonce-8510`.

In addition to the two overall measures for the perceived difference between measurement and simulation, ten perceptual qualities from the *Spatial Audio Quality Inventory* (SAQI) were selected based on informal prior listening according to their relevance and with an eye on completeness. The selection covers sound source related aspects (*source position*, *source extension*, *distance*, *localizability*), coloration (*tone color bright/dark*), the response of the acoustic environment (*duration of reverberation*, *envelopment by reverberation*), the temporal behaviour (*crispness*), and also includes the holistic measures *difference* and *clarity*. Some of the original SAQI items were combined to limit the duration of the listening test, such as *source position*, condensed from *horizontal* and *vertical direction*, and *source extension*, condensed from *depth*, *width*, and *height*. The participants

received written circumscriptions (from Lindau *et al.*[60]) and oral explanations of the qualities before the test started.

Two types of audio content were selected for SAQI testing: The pink noise pulse already used for testing authenticity was believed to best reveal artifacts for most selected qualities, and an anechoic recording of Mozart's string quartet No. 1 (bars 1-6) was taken as typical real-life content. The four tracks of the string quartet recording were assigned to the four sources arranged on stage in a semi-circular setup, and the noise pulse was played only by the rightmost source of the virtual string quartet.

Auralizations based on simulated BRIRs were compared to their measured counterparts in an interface with four continuous sliders. For orientation, the sliders had tick marks at the center, close to the end points, and at three equally spaced intermediate points of the scale. The scale labels were displayed above and below the sliders. Two buttons positioned below each slider, labeled *A*, *B*, were used to start the auralizations with *A* starting the reference and the four simulations randomly assigned to four *B* buttons. While the audio content was held constant for each rating screen, the qualities to be rated were presented in randomized order. The participants could listen as long as they needed, and switch between the four conditions on each rating screen.

Twenty-nine participants (8 female, 21 male, mean age 34 years) took part in the listening test. Twenty-four participants had already done listening tests before, 14 were experienced with room acoustical simulation, and 11 were experienced with binaural synthesis. On average, the subjects were concerned 2 hours per day with listening, playing, or working with audio. After the participants had been informed about the purpose of the experiment, the test for plausibility was conducted first, followed by the test for authenticity and the SAQI. The order of the three tests was identical for all participants, because previous exposure to the test environment should generally be avoided concerning the plausibility measure[61].

The plausibility and authenticity tests employed the medium size room only since informal prior listening had shown that the overall quality of each algorithm did not differ substantially among the three acoustic environments. For these two tests, each participant evaluated only one randomly assigned simulation algorithm, i.e. each algorithm was tested by 7 subjects (the 29th subject was discarded in this case). Each subject was presented the whole set of rooms and algorithms with varying audio content during the SAQI test.

Each test included a separate training to familiarize the participants with the interface, stimuli, and test procedure. Subjects were encouraged to move their heads and compare the auralizations at different head orientations, as this might provide additional cues. The entire test took 90 minutes on average, including general instructions, training, and short breaks between the three sections. Throughout the session, the experimenter was sitting behind a screen, not visible to the participant to avoid potential distractions. The test

[60] A. Lindau, et al. (**2014**b). "A Spatial Audio Quality Inventory (SAQI)" *Acta Acust. united Ac.*

[61] A. Lindau and S. Weinzierl (**2012**). "Assessing the plausibility of virtual acoustic environments" *Acta Acust. united Ac.*

was conducted in a quiet environment with a reverberation time of $T_m = 0.77$ s.

6.3 Results

In two sections, exemplary results for the comparison of measurements and room acoustic simulations are shown both for the simple scenes (scene 1–8), highlighting the modeling of specific acoustical phenomena, and for the complex scenes (scenes 9–11), highlighting the performance of room acoustical simulation and auralization software in real-world situations. The results are anonymized, with letters *A* to *F* assigned to the participating simulation algorithms. Only a selection of the results are discussed, while a comprehensive overview of all results including the exact source and receiver positions for every scene is given in the supplemental material (SuppPub1–3)[62]. Since some software teams contributed to selected cases only, the number of participants differs from scene to scene.

[62] See Supplementary materials at http://dx.doi.org/10.14279/depositonce-8510.

6.3.1 Simple scenes

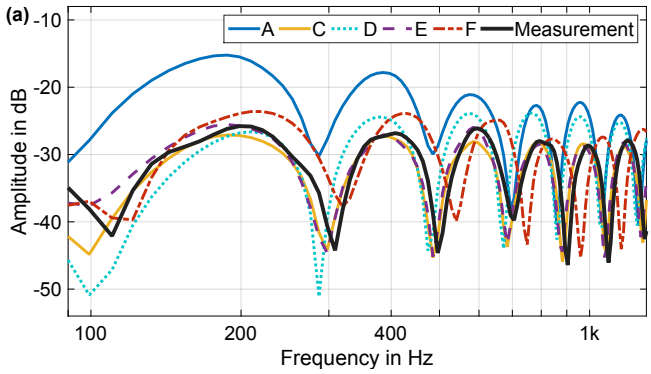
Specular reflections

Modeling a specular reflection is a simple task for an algorithm based on GA, in which case the addition of reflected energy to the direct sound results in a comb filter-like magnitude spectrum. Fig. 6.1 (a) shows the results for a reflection on a quasi infinite rigid surface (scene 1, floor of the hemi anechoic chamber) for incident and exit angles of $\gamma = 45°$. The line of sight distance between source and receiver was 4.2 m, and the source/receiver were 3 m away from the point of reflection. The comb filter effect is visible for all algorithms with small differences in the frequencies of notches and peaks due to minor deviations in the positioning of the sources/receivers between measurements and simulations. When the rigid surface is replaced by an absorber, results show that for all algorithms the comb filter effect becomes weaker for higher frequencies due to the increasing absorption (cf. SuppPub2[63], Figs. S2-3 and S2-4).

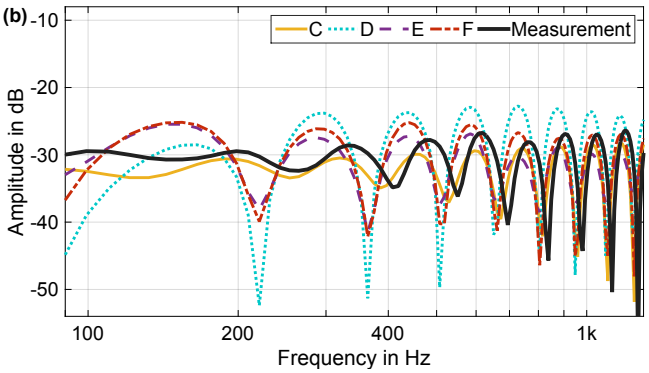
[63] See Supplementary materials at http://dx.doi.org/10.14279/depositonce-8510.

In scene 2, a reflection on a finite medium density fibreboard plate with an edge length of 1 m and 25 mm thickness was measured. Fig. 6.1(b) shows results in the frequency domain for incident and exit angles of $\gamma = 45°$. The distance between source and receiver was 5.7 m, and the source/receiver were 4 m away from the point of reflection. Due to the limited size of the reflector, most of the energy below approximately 300 Hz is diffracted around the plate and the comb-filter is less pronounced in this case. This was only correctly modeled by *C*, which includes a first-order edge diffraction model, whereas the remaining algorithms show a pronounced but "wrong" comb filter effect also for low frequencies and a largely correct simulation only for frequencies above 600 Hz. Results of the reflection on an array (scene 4, cf. SuppPub1[64], Fig. S1-17) show that for complex reflector structures, even more substantial deviations from the mea-

[64] See Supplementary materials at http://dx.doi.org/10.14279/depositonce-8510.



(a)



(b)

Figure 6.1: Specular reflections: Magnitude spectra of measured and simulated IRs for the reflection on a quasi infinite (a) and finite rigid plate (b). Both cases are for incidence/exit angles of $\gamma = 45°$ (scene 1 and 2; source position LS02; receiver position MP02, cf. SuppPub1, Figs. S1-1/2 and S1-7/8 for scene geometry).

surement can be observed in the frequency domain for all software, in particular for *D*, which failed to include a reflection at all in case of the *off center* setup (cf. SuppPub2[65], Figs. S2-23, S2-25, S2-27). In both situations shown in Fig. 6.1, software *D* shows a slightly distorted spectral shape in favor of high frequencies, whereas software *A* shows an emphasis of low frequencies for the infinite plate.

[65] See Supplementary materials at `http://dx.doi.org/10.14279/depositonce-8510`.

Diffuse reflections

To investigate in how far the algorithms can handle diffuse reflections, a one-dimensional diffusor consisting of periodically arranged wooden beams was placed on the floor of the hemi anechoic chamber (scene 1). In contrast to all other scenes, no scattering data were provided in this case. Instead, the participants were asked to model the scattering according to the demands of their software, which they did by using the geometrical diffusor model rather than assigning scattering coefficients derived from the provided dimensions of the diffusor to a single surface. Results for incident and exit angles of $\gamma = 45°$ are given in Fig. 6.2. The distance between source and receiver was 4.2 m, and the source/receiver were 3 m away from the point of reflection. No participant was able to match the measured frequency response, which can be described by an irregular comb filter. Software *C* and *F* result in a frequency response similar to the measurement, but an inaccurate temporal modeling of the diffuse reflections (cf. SuppPub2[66], Fig. S2-9) leads to a misalignment of peaks and notches in the frequency response.

[66] See Supplementary materials at `http://dx.doi.org/10.14279/depositonce-8510`.

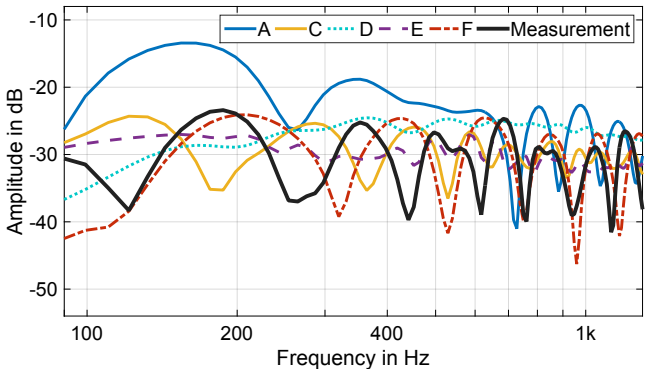


Figure 6.2: Diffuse reflections: Magnitude spectra of measured and simulated IRs for the reflection on a one-dimensional diffusor, and incidence/exit angles of $\gamma = 45°$ (scene 1; source position LS02; receiver position MP02, cf. SuppPub1, Fig. S1-20/21 for scene geometry).

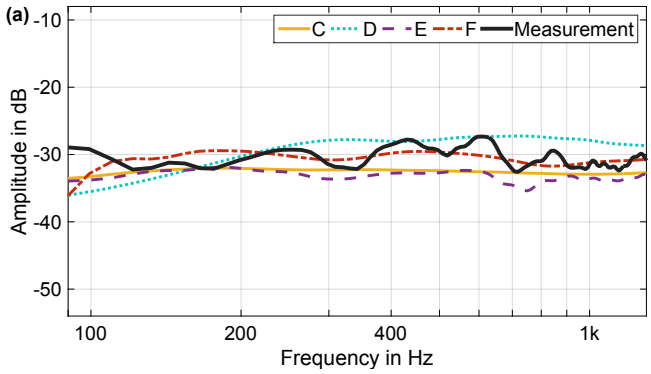
Diffraction

In case the direct sound path is close to objects or edges, diffraction adds energy to the direct sound. This causes a temporal broadening of the main impulse and/or an isolated reflection, which leads to a weak and irregular comb filter structure in the magnitude spectrum. This can be observed in the measurement depicted in Fig. 6.3(a) for a source 4 m in front and a receiver 1 m behind a medium density fibreboard panel with edge lengths of 1 m (scene 2). The source is visible from the receiver, and the direct sound path has a distance of 0.7 m to the reflector panel. While simulations from *D*, *E*, and *F* show a small extent of irregularity, *C* disregards this effect completely. Although all algorithms come relatively close the measurement, as the influence of the diffraction wave in the illuminated region is small, slightly audible coloration artifacts can be expected.

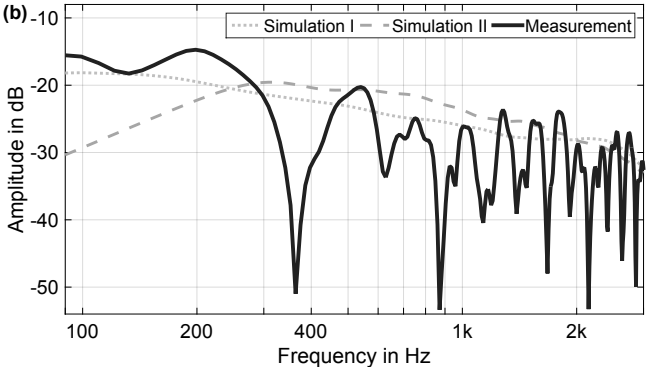
If the source is not visible to the receiver, i.e., if the direct sound path is blocked by an object, significant energetic contributions come from diffraction around objects and/or transmission through objects. Fig. 6.3(b) shows results for the diffraction around a quasi infinite medium density fibreboard wedge with a height of 2.07 m and a thickness of 25 mm (scene 5). The source and receiver were positioned 3 m in front and behind the partition at a height of 1.23 m. Because only two participants were able to simulate first-order diffraction, no letters are assigned to the simulation results in order to keep the anonymity of the participants. The results show that both programs are able to match the general trend of the measured curve where the diffracted energy arriving at the receiver decreases with increasing frequency. Apparently, the reflections on the rigid floor in front of and behind the partition, which create the comb filter structure in the measured frequency response, are not modeled. When comparing the two simulation results, a similar result can be observed for frequencies above 250 Hz while the curves substantially deviate for frequencies lower frequencies, reaching a difference of more than 10 dB for 100 Hz.

Coupled volumes

Coupled volumes, as they are used, for example, in concert hall de-



(a)



(b)

Figure 6.3: Diffraction: Magnitude spectra of measured and simulated IRs for grazing sound incidence at a finite rigid plate (a; scene 2; source position LS05; receiver position MP04, cf. SuppPub1, Figs. S1-7/8 for scene geometry) and diffraction on a quasi infinite wedge (b; scene 5, source position LS01; receiver position MP01, cf. SuppPub1, Figs. S1-20/21 for scene geometry).

sign to achieve a variable reverberation time, typically lead to a double sloped EDC[67]. Measured and simulated EDCs for the 1 kHz octave are shown in Fig. 6.4 for a reverberation chamber which was coupled to the laboratory by a door with an opening angle of $\phi = 30.4°$ (scene 8). Source and receiver were both located inside the laboratory with distances of 2.4 m and 2.2 m, respectively, to the door. The double sloped decay is clearly visible in the measured data, where the transition between decay rates of the reverberation chamber and laboratory room appears at approximately $t = 0.3$ s. When analyzing the results of this scene, it has to be considered that the EDC simulation is sensitive to the ratio of the reverberation times of both individual rooms, thus, highly depends on the provided boundary conditions. In most cases, the simulations exhibit only a weakly double sloped EDC, with the exception of *A* that seems to correctly simulate both decay rates but fails in the correct simulation of the transition time. EDCs evaluated for different octave bands and a door opening angle of $\phi = 4.1°$ show the same trends (cf. SuppPub2[68], Figs. S2-34 to S2-37).

[67] N. Xiang, et al. (**2009**). "Investigation of acoustically coupled enclosures using a diffusion-equation model" *J. Acoust. Soc. Am.*

[68] See Supplementary materials at `http://dx.doi.org/10.14279/depositonce-8510`.

6.3.2 *Complex scenes*

Room acoustical parameters

Figure 6.5 shows the reverberation time T_{20} estimated from measured and simulated RIRs and averaged across ten source and re-

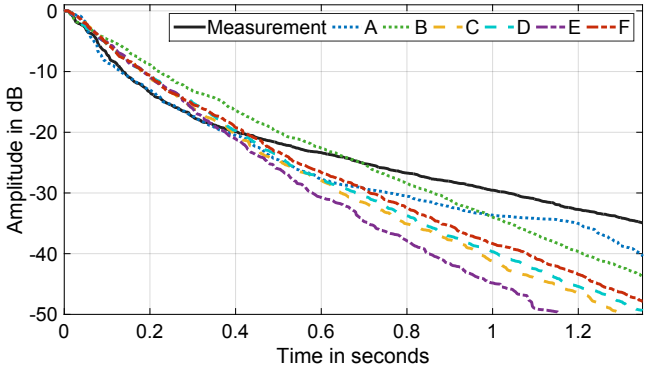


Figure 6.4: Coupled volumes: Measured and simulated energy decay curves of the coupled rooms for the 1 kHz octave band and a door opening angle $\phi = 30.4°$ (scene 8, source position LS02; receiver position MP03, cf. SuppPub1, Figs. S1-28/29 for scene geometry).

ceiver positions. Fgure 6.5 also shows the Eyring reverberation times[69] calculated based on the room volumes provided in Table 6.2.3 and the absorption coefficients provided to the software teams.

In contrast to the simple scenes, the differences between measurement and simulation here refer both to deficits of the simulation algorithms and the possibly incorrect estimation of absorption coefficients with the in-situ measurements conducted. Both uncertainties also occur in room acoustical design practice; the results are thus a valid indication of reliability of room acoustical simulation as a planning and design tool. As a result of both sources of error, a trend for overestimating the actual reverberation times at low frequencies and underestimating them at high frequencies can be observed. The simulations resulted in reverberation times that are closer to the Eyring estimates than to the measured values in most cases. Because the simulations and the Eyring estimates are based on the provided absorption data, this might indicate that differences between measurements and simulations are dominated by uncertainty in the absorption coefficients. The differences between measurement and simulation are particularly high for the 125 Hz and 250 Hz octave bands, where the measured reverberation times are, on average, overestimated by 58% (125 Hz) and 35% (250 Hz). For the mid-frequency range (500 Hz – 2 kHz), there is not systematic deviation; the differences between simulation and measurement are, however, still above the JND in most cases. A systematical overestimation of the absorption coefficients at 1 kHz, which was observed in RR-I[70], does not appear in the three scenes tested here.

Analyzing the differences across simulation results, it is apparent that the variance decreases with increasing frequency: While the T_{20} values exhibit a range of about 1 s at 125 Hz in all three rooms, a range of approximately 0.5 s can be observed at 4 kHz if neglecting single outliers. In addition, the differences between simulation algorithms are room dependent to a surprisingly large degree. For instance, algorithms *B* and *E* exhibit reverberation times around 1.6 s at 125 Hz for the small room, while the remaining results cluster around 2.5 s. For the medium and large room, however, *A* and *F* produce reverberation times that are about 0.8 s higher than those from the remaining algorithms. Moreover, a closer look at results

[69] H. Kuttruff (**2009**). *Room acoustics*.

[70] M. Vorländer (**1995**). "International round robin on room acoustical computer simulations" in *15th International Congress on Acoustics*.

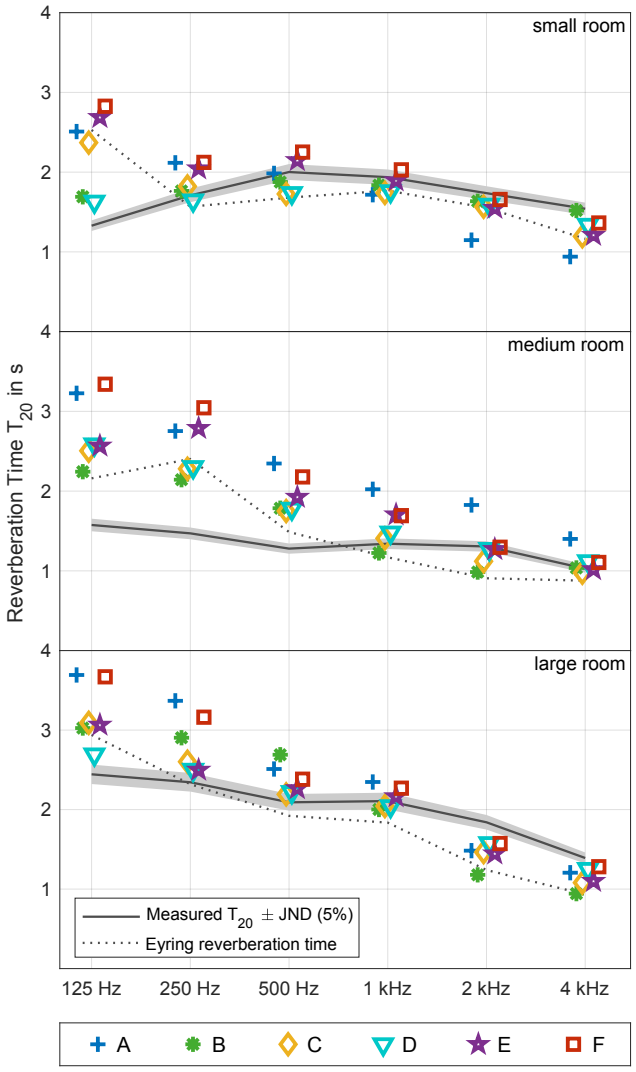


Figure 6.5: Reverberation time: T_{20} calculated from measured and simulated RIRs. Results are averaged across ten source/receiver positions and evaluated for six octave bands. To improve the readability, the simulation results are shifted in horizontal direction, and the reference values are connected by lines.

from A shows that the T_{20} for $f \leq 2$ kHz is below the average in the small room, above the average in the medium room, and clusters with the remaining algorithms in the large room.

The results for additional room acoustical parameters and all source and receiver positions show similar trends (cf. SuppPub3[71], Figs. S3-1 to S3-31). While the EDT is overestimated at low frequencies and underestimated at high frequencies, the opposite holds for the clarity (C_{80}) and definition (D_{50}). Room acoustic parameters for individual source/receiver positions were analyzed for 1 kHz. The correlation between values based on measurements and simulations, however, was statistically non-significant with the exception of C_{80} and D_{50} where correlations between 0.7 and 0.9 were observed for the large room and all algorithms. For the EDT, a correlation of -0.7 occurred for D in the small room, indicating a reversed spatial dependency in this case (cf. SuppPub3[72], Figs. S3-1 to S3-2).

[71] See Supplementary materials at `http://dx.doi.org/10.14279/depositonce-8510`.

[72] See Supplementary materials at `http://dx.doi.org/10.14279/depositonce-8510`.

Spectral differences

Spectral differences between measured and simulated BRIRs with

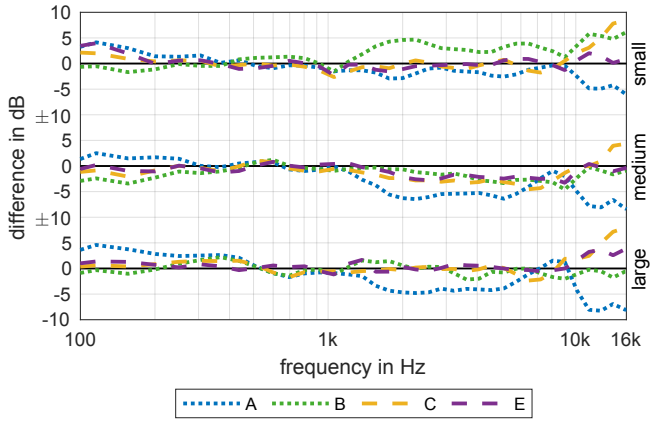


Figure 6.6: (Color online) Energetic differences between simulated and measured BRIRs in auditory filter bands averaged across source positions and ears for the three complex rooms (cf. SuppPub1, Figs. S1-15 – S1-17 for scene geometry).

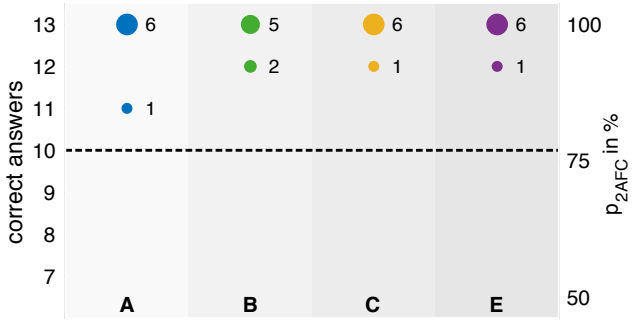


Figure 6.7: Results of the test for authenticity: Numbers of correct answers (left y-axis) and corresponding detection rates in percent (right y-axis). The size of the dots and the numbers next to them show how many participants had identical results. Results on or above the dashed line indicate significant differences, i.e., non-authentic simulations. 50% correct answers denote guessing, and 75% the threshold of perception.

neutral head orientation are shown in Fig. 6.6. Averaged results are given because correlations among ears and source positions were high (Ø = 0.8), although average errors are between 2 dB and 2.5 dB. Software *A* always shows a bass boost and a lack of energy at high frequencies, while software *C* exhibits a high frequency boost in all cases. Overall, smallest differences were observed for *E* (Ø = 1.3 dB), followed by *B* and *C* (Ø = 2.2 dB), and *A* (Ø = 3.2 dB).

Perceptual evaluation

Taking into account the previously analyzed differences in T_{20} and the magnitude spectra, *authenticity* of the simulated rooms can presumably not be reached. This is proved by the results from the test for authenticity, assessing the perceptual identity of measured and simulated BRIRs of the medium room (scene 10) in a 2AFC listening test paradigm (cf. Fig. 6.7). Apparently, all participants could reliably identify differences between reality and simulations, with detection rates of $p_{2\mathrm{AFC}} \geq 0.98$, and the number of correct answers clearly exceeding the critical value of $N_{\mathrm{crit.}} = 10$ for all simulation algorithms. Thus, none of the auralizations managed to be indistinguishable from the measured reference. This means that at the time being, blind simulations starting without a priori knowledge about reverberation times, etc., cannot lead to authentic results.

Results for the evaluation of *plausibility*, testing the credibility of simulations vs. measurements with respect to an inner acoustic reference, are given in Fig. 6.8. The simulations were perceived as

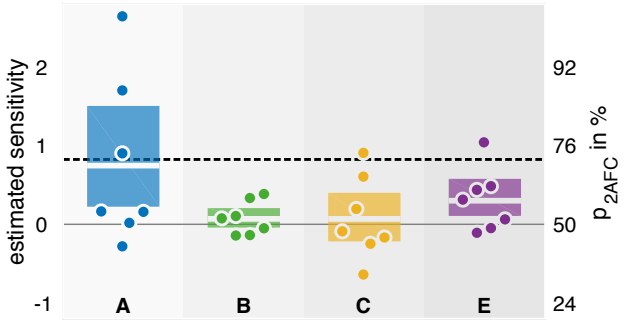


Figure 6.8: Results of the test for plausibility: Estimated individual sensitivities $\hat{d}'$ (left y-axis), and corresponding 2AFC detection rates $p_{2\text{AFC}}$ (right y-axis) are given by the points (offset in horizontal direction to improve readability). Individual sensitivities on or above the dashed line indicate non-plausible simulations. 50% correct answers denote guessing, and 75% the threshold of perception. The boxes show the group mean and 90% bootstrapped confidence intervals (non-parametric resampling, bias corrected and accelerated CI calculation. A tabular overview of the individual results is given in SuppPup3, Fig. S3-35.

plausible in most cases, indicated by sensitivity values below the critical value. However, slight differences between the algorithms emerge: Simulation *B* was perceived as plausible by all participants ($\hat{d}'_{\text{mean}} = 0.07$), one participant detected artifacts in simulations *C* and *E* ($\hat{d}'_{\text{mean}} = 0.07$, and 0.3), and simulation *A* was perceived as implausible by 3 participants ($\hat{d}'_{\text{mean}} = 0.75$).

Differences in *specific auditory qualities* were measured using selected attributes of the SAQI (Fig. 6.9). Median values and 95% bootstrap confidence intervals (CIs: non-parametric resampling, bias corrected and accelerated CI calculation[73]) are given because the ratings were not normally distributed in the majority of cases. The auralizations of simulated BRIRs were directly compared to their measured counterparts, thus, a rating of 0 indicates no perceivable difference, and a rating of ± 1 stands for maximum differences.

The cases where the CIs do not overlap zero are taken as an indication of significant deviations between measurement and simulation (cf. Fig. 6.10). Here, differences become obvious between the different algorithms, between the two audio contents and, to a lesser degree, between the three different rooms. Whereas the softwares *B*, *C*, and *E* show significant deviations in 12% to 30% of the experimental trials, this is the case in 88% of the cases for *A*. Remarkably, no large perceptible deviations from reality were observed for *E* and the string quartet in the small and medium room. Considering only *B*, *C*, and *E*, large deviations were mainly observed for *difference*, *tone color bright/dark*, and *source position*. Moreover, no major differences were observed for *envelopment*. Also visible is a large difference between the two audio contents with the pulsed pink noise making the differences in most qualities more noticeable.

These visual inspection observations are confirmed by a three-factorial analysis of variance (ANOVA for repeated measurements) to test for significant differences concerning the normally distributed rating item *difference* with the factors *algorithm* (A, B, C, E), *room* (small, medium, large), and *content* (music, noise). It shows a highly significant main effect for the factor *algorithm* ($F(3) = 128.9$, $p < 0.001$, $\eta_p^2 = 0.82$). Bonferroni-corrected post-hoc t-tests showed that *E* performed significantly better than the remaining algorithms (all $p < .001$), while *A* was significantly worse than the others (all $p < .001$). Variation of the presented *room* led to a small but significant

[73] J. Carpenter and J. Bithell (**2000**). "Bootstrap confidence invervals: when, which, what? A practical guide for medical statisticians" *Statistics in Medicine*.

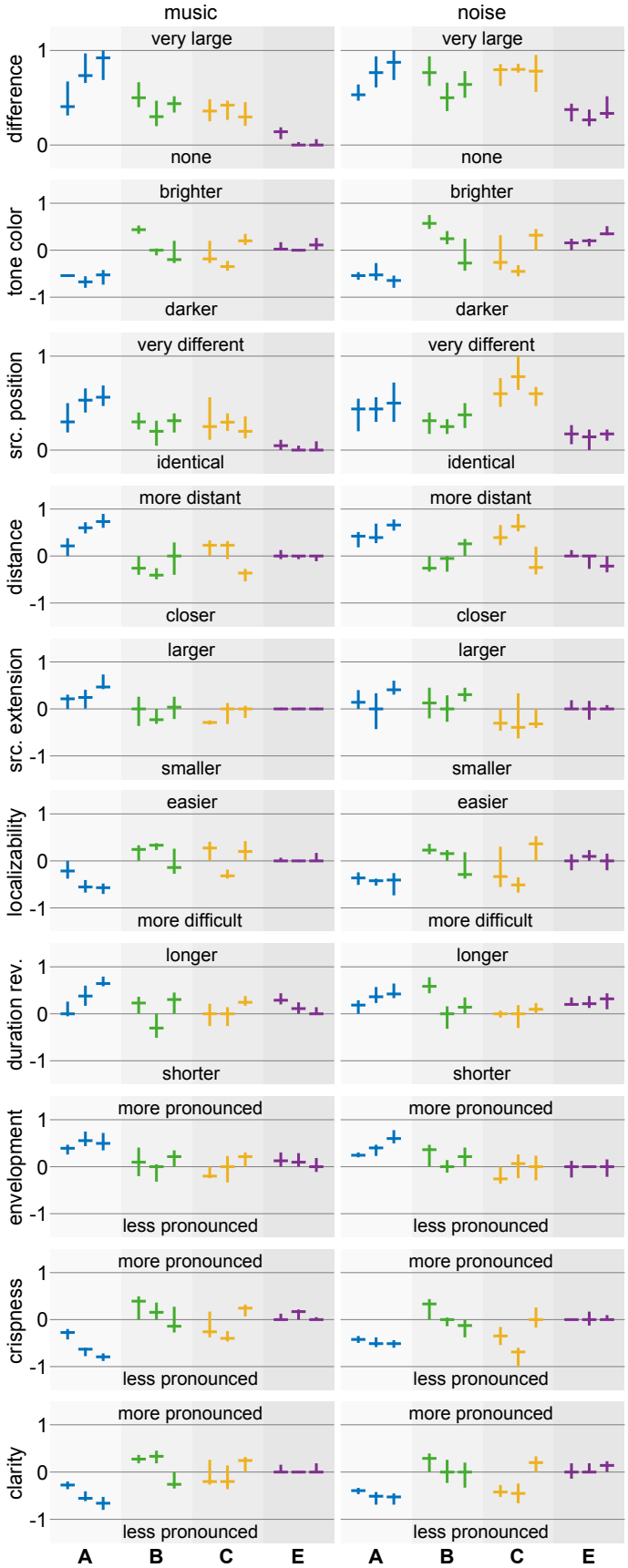


Figure 6.9: Differences in specific auditory qualities, measured with attributes of the Spatial Audio Quality Inventory (SAQI), showing the median of differences between simulation and measured reference (horizontal lines) with 95% bootstrap confidence intervals (vertical lines). The ratings were given for music (string quartet, left) and pulsed pink noise (right) as audio content, and for the small, medium and large room (from left to right).

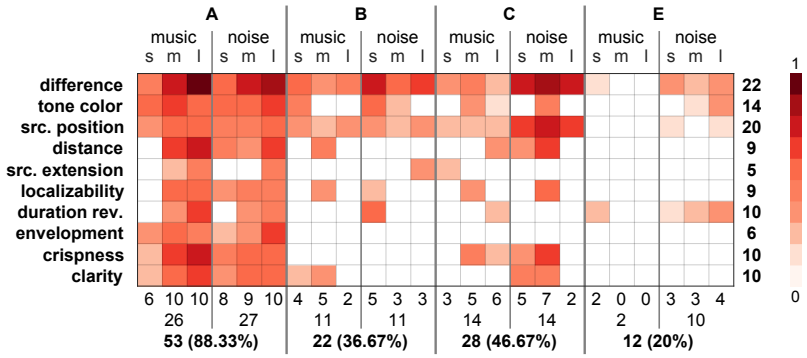


Figure 6.10: Results of the SAQI test: Degree of deviations by algorithm, audio content, room-size and perceptual quality: White areas denote CIs overlapping with 0, shaded areas denote CIs *not* overlapping with 0, in which case the shading denotes the absolute median ratings in the range between 0 and 1 as indicated by the color bar. Numbers indicate the sum of significant deviations across rows and columns. Results for the small, medium, and large room are indicated by the letters *s*, *m*, and *l*.

main effect ($F(2) = 3.2$, $p = 0.048$, $\eta_p^2 = 0.1$), with larger perceived deviations from the reference for the large room compared to the other two (estimated marginal means: $\epsilon_{small} = 0.49$, $\epsilon_{medium} = 0.48$, $\epsilon_{large} = 0.52$; SE's $\approx .025$). Moreover, perceived differences turned out to be significantly larger for pink noise compared to the musical content across all algorithms and rooms ($F(1) = 78$, $p < 0.01$, $\eta_p^2 = 0.74$). A detailed report of the ANOVA statistics is given in SuppPub3[74], Figs. S3-36 to S3-39.

To highlight the qualitative pattern of perceptual differences between simulation and measurement, a three-way multivariate analysis of variance (MANOVA for repeated measurements) was carried out for all attributes except *difference*. An inspection of the model residuals proved that the requirement of normality was met (SuppPub3[75], Fig. S3-40). Here, the factor *content* had a multivariate main effect (*Pillai's Trace* $= .753$, $F(9, 20) = 6.79$, $p < .001$, $\eta_p^2 = 0.75$), with always larger perceived deviations for the noise signal, although not every univariate main effect is significant. The factor *algorithm* also generated a multivariate main effect (*Pillai's Trace* $= 1.715$, $F(27, 234) = 11.572$, $p < .001$, $\eta_p^2 = 0.57$) with significant univariate main effects for *all* qualities (all $p < .01$). Finally, a multivariate main effect (*Pillai's Trace* $= .952$, $F(18, 98) = 4.944$, $p < .001$, $\eta_p^2 = 0.48$), encompassing five significant univariate main effects was occurring for the factor *room* (all $p < .01$). Noteworthy, the factor *algorithm* explains considerably more variance than the *room* ($\eta_p^2 = 0.57$ vs. $\eta_p^2 = 0.48$), and also causes the largest range in the estimated marginal means ($\mu(\Delta\epsilon) = 0.47$ vs. 0.1 *room*, and 0.07 *content*, $\mu(\cdot) :=$ average across qualities) showing that the *algorithm* has the strongest influence on the perceived differences between simulations and reference. The interactions *algorithm* $\times$ *content*, and *algorithm* $\times$ *room* are significant for all qualities as well (*Pillai's Trace* $= .908$, $F(27, 234) = 3.763$, $p < .001$, $\eta_p^2 = 0.3$, and *Pillai's Trace* $= 1.504$, $F(54, 990) = 6.133$, $p < .001$, $\eta_p^2 = 0.25$), demonstrating that no single algorithm clearly outperforms the others with respect to all rooms, content types, and perceptual qualities.

In the following, we restrict the presentation of the univariate results to the qualities *tone color*, *source position*, and *localizability* be-

[74] See Supplementary materials at `http://dx.doi.org/10.14279/depositonce-8510`.

[75] See Supplementary materials at `http://dx.doi.org/10.14279/depositonce-8510`.

Quality	A	B	C	E
difference	0.68 (0.29)	0.53 (0.19)	0.56 (0.06)	0.24 (0.10)
tone color	-0.54 (0.13)	0.18 (0.62)	-0.05 (0.56)	0.15 (0.18)
src. position	0.48 (0.16)	0.32 (0.11)	0.47 (0.07)	0.14 (0.06)
distance	0.46 (0.43)	-0.12 (0.28)	0.13 (0.56)	-0.07 (0.17)
src. extension	0.21 (0.33)	0.06 (0.28)	-0.12 (0.09)	0.03 (0.06)
localizability	-0.37 (0.16)	0.10 (0.25)	-0.01 (0.50)	0.07 (0.13)
duration rev.	0.34 (0.41)	0.09 (0.54)	0.05 (0.13)	0.20 (0.04)
envelopment	0.40 (0.29)	0.06 (0.28)	-0.03 (0.11)	0.04 (0.14)
crispness	-0.48 (0.26)	0.07 (0.24)	-0.20 (0.62)	0.08 (0.05)
clarity	-0.44 (0.17)	0.07 (0.28)	-0.12 (0.41)	0.04 (0.06)
∅	0.44 (0.26)	0.16 (0.31)	0.17 (0.31)	0.11 (0.10)

Table 6.3: Estimated marginal means ϵ of perceived differences between measurement and simulation for ten perceptual qualities (SAQI attributes), and their range $\Delta\epsilon$ across rooms (in parentheses). The marginal means were obtained by analysis of variance. The last row shows the mean absolute values.

cause they show the largest differences in general (cf. Fig 6.10), and also made the strongest contribution to the previously described mixed regression model (cf. Table 6.3.2). First of all, simulations from *A* are considerably darker than the reference (estimated marginal mean $\epsilon = -0.54$), whereas *B* and *E* sound somewhat brighter ($0.15 \leq \epsilon \leq 0.18$). Only *C* sounded neutral on average ($\epsilon = -0.05$). While the results for *A* and *E* were almost constant across rooms ($\Delta\epsilon = 0.18$), the room influence was strong for *B* and *C* ($\Delta\epsilon \approx 0.59$). This can also be observed for the *localizability* ($\Delta\epsilon \approx 0.15$ vs. $0.25 \leq \Delta\epsilon \leq 0.5$). The *source position* was relatively accurate for *E* ($\epsilon = 0.14$), and less accurate for the remaining algorithms ($0.32 \leq \epsilon \leq 0.48$). However, the univariate room effect had been statistically non-significant in this case ($\Delta\epsilon < 0.16$).

To give an overview of the size of the simulation-related deviations in the various qualities, the estimated marginal means for the software algorithms and their variation across rooms are given in Table 6.3.2. The overall picture is in line with previous observations: Smallest differences were observed for *E*, medium differences for *B* and *C*, and largest differences for *A*. These differences are quite consistent across the three rooms. A detailed report of the MANOVA statistics is given in SuppPub3[76], Figs. S3-40 to S3-51.

Finally, a mixed regression model[77] was estimated with *difference* as dependent and the other nine qualities as independent variables. This was done to assess the importance of each perceptual dimension for the degree of overall perceived differences as expressed in the *difference* score. For this purpose, absolute values were taken, thus assuming that positive and negative deviations (e.g. *tone color*: darker – brighter) would equally contribute to the perceived *difference*. To test for multicollinearity, bivariate Pearson correlations between absolute scores of ratings of all attributes were calculated, with $r = 0.24$ on average and $r \leq 0.51$ in all cases. Thus, no qualities were excluded from the regression model and only removed in case of non-significant contributions to the prediction. The model included a random intercept term for *participant* in order to control for individual rating thresholds, and assumed a first-order auto-regressive residual covariance matrix due to repeated measurements. An inspection of the

[76] See Supplementary materials at `http://dx.doi.org/10.14279/depositonce-8510`.

[77] J. J. Hox (**2010**). *Multilevel Analysis. Techniques and Apllications*.

Quality	Beta weight
tone color	0.294
src. position	0.162
localizability	0.139
clarity	0.084
distance	0.083
crispness	0.082
envelopment	0.081

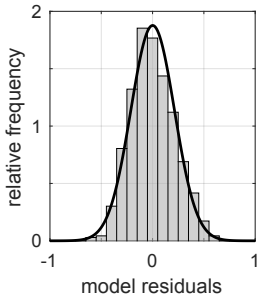


Table 6.4: Mixed regression model showing the influence of the different qualities on the perceived overall *difference*. Left: Standardized model estimates (beta weights). All included qualities have significant contributions with $p < 0.05$. Right: Distribution of the model residuals and the corresponding normal probability density function with identical mean and standard deviation.

model residuals showed that the requirement of normality was met (cf. Table. 6.3.2). The final model (with *duration of reverb* and *crispness* removed due to non-significant influence) accounts for $R^2 = 55.9\%$ of the variance (marginal $R^2 = 41.3\%$[78]) and is shown in Table 6.3.2. *Tone color* has the largest influence on the *difference*, followed by *source position* and *localizability*.

[78] S. Nakagawa and H. Schielzeth (**2013**). "A general and simple method for obtaining R^2 from generalized linear mixed-effects models" *Methods in Ecology and Evolution*.

6.3.3 *Primary research data*

The database of acoustical scenes (Table 6.2.1), including all data provided to the participants (3D models, absorption and scattering coefficients, source and receiver information) is available as an electronic publication[79,80]. It now contains also the reference measurements that were not available for the participants. A description of all scene configurations and a comprehensive compilation of all results of the physical and perceptual evaluation is available in the supplemental material[81]. These also include examples of the audio stimuli of the listening tests, with a static version of the originally dynamic binaural auralizations for neutral head orientation. Auralizations based on the measured data are directly followed by auralizations based on the simulated data as verbally announced. The audio files are diffuse field compensated and should be played back via headphones.

[79] L. Aspöck, et al. (**2018**). *GRAS – Ground Truth for Room Acoustical Simulation* `https://dx.doi.org/10.14279/depositonce-6726`.

[80] Please note that an updated version named BRAS has been published in the meantime: L. Aspöck, et al. (**2019**). *BRAS – A Benchmark for Room Acoustical Simulation* `https://dx.doi.org/10.14279/depositonce-6726.2`.

[81] See Supplementary materials at `http://dx.doi.org/10.14279/depositonce-8510`.

6.4 *Discussion*

When interpreting the present results, readers should be reminded that the participating software developers were not allowed to change the input data (source/receiver directivities and absorption/scattering coefficients) with the exception of the geometrical resolution of the 3D models. This *blind* evaluation was considered the best way to assess quality differences between simulation algorithms because the fitting of parameters would compensate for shortcomings of the numerical simulation and make deviations from the reference – and most likely also between algorithms – appear smaller than they really are[82].

In the *physical* evaluation of the participating algorithms, simulated IRs were compared with the measured reference. Only in the case of specular reflections on a quasi infinite surface (scene 1), the algorithms were able to match the spectral and temporal behaviour of the reference well. Small deviations already occurred for reflec-

[82] M. Vorländer (**2013**). "Computer simulations in room acoustics: Concepts and uncertainties" *J. Acoust. Soc. Am.*

tions on a quasi infinite two-dimensional diffusor (scene 1), where no algorithm was able to exactly match the comb filter structure of the magnitude spectrum. Even if this might have only minor consequences for the room acoustical parameters of complex environments[83,84], it is likely to introduce coloration in the modeling of early reflections. For the reflection on the finite plate (scene 2), only one algorithm with an implementation of first-order edge diffraction was able to approximate the diffraction around the plate for wave lengths, which are large compared to the dimension of the plate. No algorithm, however, accurately modeled the coloration due to grazing sound incidence at the same plate. Because almost all relevant acoustic environments contain reflectors or objects of limited size, this is another source that might introduce severe coloration in modeling early reflections. Whereas the precise modeling of scattering effects of diffusing structures will remain a challenge in GA, there are approaches to account for diffraction in image source models and ray tracing[85], which the authors believe deserve more attention to improve the tested algorithms. Noteworthy, this was already mentioned during the first round robin on room acoustical computer simulation in 1995[86]. Modeling diffraction becomes even more important in case the direct sound path is blocked by an object. Results of scene 5 showed that by modeling only a single diffraction path the spectral shape of the transfer function is not well preserved. Neglecting diffraction entirely will be even worse, keeping in mind the increased relevance of dynamic simulations for virtual acoustic reality, where sudden jumps in loudness, tone color, or source position are likely to be perceived if the listener passes objects blocking the direct sound path. The double sloped decay of a coupled room (scene 8) could also not be modeled by most simulation algorithms, which might cause differences in the perceived reverberation tail. The observed errors could be caused by an insufficient number of rays in the ray tracing and/or, again, by omitting diffraction around the area that couples the two volumes. At this point, analytical and stochastic models for the energy decay of coupled volumes (cf. Luizard *et al.*[87] for an overview) could be used as a reference to improve the behavioor of GA-based simulation algorithms.

The simple scenes discussed above (scenes 1–8) could be accurately described by means of a 3D model, source and receiver directivities, as well as absorption and scattering coefficients or complex impedances. For complex scenarios, however, the acoustic surface properties (absorption, scattering) were estimated based on narrow band in situ measurements[88] or from material descriptions and pictures. As a consequence, differences between reference and simulation may either stem from shortcomings of the algorithms or uncertainties in the description of the boundary conditions. Since this is the case also in real-life applications, the discussion of the results for scenes 9 – 11 gives an impression of the general reliability and the systematic errors that occur in the application of these algorithms for acoustic planning tasks.

[83] I. Bork (**2005**a). "Report on the 3rd round robin on room acoustical computer simulation - Part I: Measurements" *Acta Acust. united Ac.*

[84] I. Bork (**2005**b). "Report on the 3rd round robin on room acoustical computer simulation - Part II: Calculations" *Acta Acust. united Ac.*

[85] L. Savioja and U. P. Svensson (**2015**). "Overview of geometrical room acoustic modeling techniques" *J. Acoust. Soc. Am.*

[86] M. Vorländer (**1995**). "International round robin on room acoustical computer simulations" in *15th International Congress on Acoustics*.

[87] P. Luizard, et al. (**2014**). "Sound energy decay in coupled spaces using a parametric analytical solution of a diffusion equation" *J. Acoust. Soc. Am.*

[88] E. Brandão, et al. (**2015**). "A review of the in situ impedance and sound absorption measurement techniques" *Acta Acust. united Ac.*

A comparison of the temporal structure of the simulated and measured IRs for the three rooms (SuppPub3[89], Figs. S3-12, S3-22, S3-32) reveals that not all strong individual reflections are correctly modeled. This is due to the missing or insufficient representation of diffraction phenomena, possibly in combination with the impact of angle dependent surface properties, which are not considered by any of the algorithms. At least for reflections that occur before the perceptual mixing time[90], the difference between the measured and simulated reflection patterns is likely to be audible.

Considering the calculated room acoustical parameters according to ISO 3382-1[91], there is no systematic deviation between measurement and simulation for values in the medium frequency range (500 Hz – 2 kHz); in many cases, the deviation is within the JND, which can be considered as a critical perceptual threshold. While a systematic overestimation of the reverberation time at 1 kHz was, unlike in the pasta[92], no longer observed, probably due to the better in-situ measurements or due to improved databases with tabulated absorption coefficients, there is still a systematic overestimation of low-frequency reverberation as well as a tendency to underestimate the reverberation time above 2 kHz. Both effects can be traced back to inaccurate absorption coefficients in connection with the geometry used, whose resolution seems to be optimal only for the middle frequency range[93].

In addition to the physical evaluation, a *perceptual* evaluation of the different simulation algorithms was conducted, based on two overall measures for the degree of perceived difference between simulation and reference, and on a qualitative description of the differences, based on attributes from the SAQI. The test for *authenticity* showed that differences between simulations under test and the reference were always audible. This finding corresponds to the physical evaluation where no algorithm met the investigated room acoustical parameters within the tolerance of the JND in all frequency bands. Considering the high sensitivity of the test and human auditory system in general, it seems unlikely that the simulation of a complex acoustic environment will be able to achieve authenticity in a blind comparison in the foreseeable future. Even if fitting the input data would be allowed it could be argued that, on the one hand, the accuracy of wave-based simulations that numerically solve the wave equation will always be limited by the quality of input data describing the sound source and the boundary conditions. This will remain a challenging and, currently, at least partially unsolved task[94]. On the other hand, the limited accuracy of the modeling of diffraction and scattering in GA was shown to introduce errors that are very likely to be audible, even if providing accurate input data.

While authenticity, denoting that the simulation sounds exactly like the real room, is a very strict criterion, *plausibility*, denoting the occurrence of obvious artifacts in the simulation which can be detected even without comparison with an explicit reference, can be considered a minimum quality criterion for virtual acoustic realities.

[89] See Supplementary materials at `http://dx.doi.org/10.14279/depositonce-8510`.

[90] A. Lindau, et al. (**2012**). "Perceptual evaluation of model- and signal-based predictors of the mixing time in binaural room impulse responses" *J. Audio Eng. Soc.*

[91] ISO 3382-1 (**2009**). *Measurement of room acoustic parameters – Part 1: Performance spaces*.

[92] M. Vorländer (**1995**). "International round robin on room acoustical computer simulations" in *15th International Congress on Acoustics*.

[93] M. Vorländer (**2013**). "Computer simulations in room acoustics: Concepts and uncertainties" *J. Acoust. Soc. Am.*

[94] M. Vorländer (**2013**). "Computer simulations in room acoustics: Concepts and uncertainties" *J. Acoust. Soc. Am.*

In the present test, three out of four algorithms were able to provide plausible auralizations for most of the participants, and only one algorithm produced detection rates well above the threshold of perception. The averaged detection rates of $0.52 \leq \overline{p}_{2\mathrm{AFC}} \leq 0.58$ for *B*, *C*, and *E* are comparable to those found for non-individual binaural simulation based on measured BRIRs ($\overline{p}_{2\mathrm{AFC}} = 0.51$[95], and $\overline{p}_{2\mathrm{AFC}} = 0.55$[96]).

The perceived *difference* between reference and simulation was mostly caused by differences in *tone color* and perceived *source position*, as could be demonstrated by regression analysis of the ratings of the SAQI attributes (Table 6.3.2). The deviations in *tone color* can be attributed to the inadequate modeling of early and late reflections for the reasons discussed above. Interestingly, the systematic low-frequency overestimation and high-frequency underestimation of T_{20} in the simulated IRs did not lead to a corresponding rating of tone color (cf. Figs. 6.5 and 6.9). In fact, the majority of the simulations by algorithms *B*, *C*, and *E* were rated as brighter than the measurement, indicating that the bass ratio, i.e., the ratio of reverberation times at low and medium frequencies, is not a reliable indicator for tone color in this case. It seems that the missing or insufficient modeling of diffraction for early reflections is what leads to an unnaturally bright sound impression. This was only not the case for software *A*, which showed a strong low-frequency boost already for the single reflection on the quasi infinite surface (Fig. 6.1a).

Differences in *source position* are most probably a by-product of spectral differences leading to mislocalizations in elevation[97,98]. Although it was not distinguished between localization errors in horizontal and vertical direction in the listening test, a vertical displacement is much more likely because the low-frequency ITD, which dominates horizontal localization[99], was well preserved in the simulations (except for *C*), which were controlled by a signal-related analysis (cf. SuppPub3[100] Fig. S3-33).

Compared to previous attempts, the current round robin has given a much more detailed insight into the performance of room acoustical simulation algorithms. This was made possible by the creation of a database of acoustical scenes with well controlled information on geometry and boundary conditions, highlighting different acoustic phenomena, and by conducting a technical *and* perceptual evaluation of the generated auralizations. This procedure entailed a larger effort on the side of the developers to calculate the required IRs, and is one reason why the number of participating software teams was lower than in previous attempts[101,102,103,104]. Some of the features that would allow easier accessibility for benchmarking tasks like the current one, however, would also be valuable extensions of the software packages for practical application. These include interfaces to import external HRTF sets into the software, the scripting and automation for different source and receiver combinations, different project variants, or different HRTF orientations for dynamic binaural synthesis. It should also be noted that the computation times for

[95] A. Lindau and S. Weinzierl (**2012**). "Assessing the plausibility of virtual acoustic environments" *Acta Acust. united Ac.*

[96] C. Pike, et al. (**2014**). "Assessing the plausibility of non-individualised dynamic binaural synthesis in a small room" in *AES 55th International Conference*.

[97] J. Blauert (**1997**). *Spatial Hearing. The psychophysics of human sound localization.*

[98] R. Baumgartner, et al. (**2014**). "Modeling sound-source localization in sagittal planes for human listeners" *J. Acoust. Soc. Am.*

[99] F. L. Wightman and D. J. Kistler (**1992**). "The dominant role of low-frequency interaural time differences in sound localization" *J. Acoust. Soc. Am.*

[100] See Supplementary materials at `http://dx.doi.org/10.14279/depositonce-8510`.

[101] M. Vorländer (**1995**). "International round robin on room acoustical computer simulations" in *15th International Congress on Acoustics*.

[102] I. Bork (**2000**). "A comparison of room simulation software – The 2nd round robin on room acoustical computer simulation" *Acta Acust. united Ac.*

[103] I. Bork (**2005**a). "Report on the 3rd round robin on room acoustical computer simulation - Part I: Measurements" *Acta Acust. united Ac.*

[104] I. Bork (**2005**b). "Report on the 3rd round robin on room acoustical computer simulation - Part II: Calculations" *Acta Acust. united Ac.*

simulating an IR strongly vary with the software, which indicates opportunities for performance optimization in some cases. Such an optimization might be a prerequisite for the future implementation of computationally more demanding models for diffraction and scattering.

Since the database and the reference measurements are now open for free access, they can also be used by the developers of room acoustical simulation software themselves to evaluate the performance of new modeling approaches. The improvement of simulation software, as well as the extension of the database by further acoustic scenes will be a reason for the authors to repeat this test in the future.

6.5 *Conclusion*

In the first round robin on room acoustical simulation and auralization, the simulation results for six simple scenes and three complex rooms provided by six teams using five different acoustic simulation algorithms were compared against measured data with respect to physical and perceptual properties. The results demonstrate that most present simulation algorithms based on GA generate obvious model errors once the assumptions of an infinite reflective baffle are no longer met. As a consequence, they are neither able to provide an exact pattern of early reflections, nor do they provide an exact prediction of room acoustic parameters outside a medium frequency range of 500 Hz–2 kHz.

In the perceptual domain, the algorithms under test could generate mostly plausible but not authentic auralizations. That means the difference between simulated and measured IRs of the same scene was always clearly audible. Most relevant for this perceptual difference are deviations in tone color and source position between measurement and simulation which to a large extent can be traced back to errors in the simulation of early reflections, due to the simplified use of random incidence absorption and scattering coefficients and the missing or insufficient modeling of diffraction. Hence, room acoustical simulations are, unlike measurement-based auralizations[105], not yet suitable to accurately predict the perceptual properties of sound sources in virtual acoustic environments at the current state of the art. Moreover, significant differences between different simulation algorithms have to be expected.

These conclusions hold for the conducted blind comparison task with initial parameter estimates, as is the case in the acoustic design of not yet existing venues. As soon as this estimate can be fitted to the measurement of a (partially) existing environment, modeling errors will become smaller automatically.

From a methodological point of view, we are convinced that the combination of an open database containing acoustic scenes[106] and a repeated comparison of different simulation algorithms against this reference could provide good prerequisites for the further improve-

[105] F. Brinkmann, et al. (**2017**a). "On the authenticity of individual dynamic binaural synthesis" *J. Acoust. Soc. Am.*

[106] L. Aspöck, et al. (**2019**). *BRAS – A Benchmark for Room Acoustical Simulation* https://dx.doi.org/10.14279/depositonce-6726.2.

ment of room acoustical simulation. Since room acoustical simulation will be more and more important for the generation of virtual acoustic realities, this evaluation should be based on physical as well as on perceptual criteria.

7
Conclusion

This thesis detailed the preparation, conduction, and evaluation of an inter-algorithm round robin on room acoustical simulation and auralization. Before implications of the results and future perspectives are discussed, a brief list of the original achievements is given.

7.1 Original achievements

Chapter 2 proofed that individual dynamic binaural synthesis based on direct acoustic measurements is indistinguishable from the real sound filed if considering audio content that is encountered in real life. While similar experiments were conducted for static individual synthesis, this is the first empirical data on the authenticity of dynamic reproduction – which is technically more demanding – including the influence of different degrees of reverberation. The results thus proved that binaural synthesis can be used as a reference system to replace cumbersome in situ experiments.

Chapter 3 showed that the effect of the torso on head-related impulse responses is audible even for source positions that cause weak torso effects. The presented study was the first investigation exceeding a mere physical description of the torso effect, including a multitude of source positions and head-above-torso orientations.

Chapter 3 also provided experimental evidence that the torso effect can be interpolated perceptually transparent from a relatively sparse dataset where the head-above-torso orientation is sampled in intervals of ten degree. This was achieved by a novel interpolation approach using to head-related transfer functions with identical head-to-source orientation and different torso-to-source orientations, thus avoiding an interpolation of the high frequency pinnae cues that underly a higher spatial fluctuation than torso cues.

Chapter 4 detailed the acquisition of the first publicly available head-realted transfer function database including multiple head-above-torso orientations. The database was moreover among the first to provide full-spherical data with a high spatial resolution for

the entire audible bandwidth, thus enabling a perceptually transparent and spatially continuous representation of the source position and head-above-torso orientation by means of interpolation. Chapters 3 and 4 laid the ground work for high quality dynamic model based auralizations of simulated binaural room impulse responses.

Chapter 5 documented the conceptual design and acquisition of the first publicly available database with a multitude of simple and complex acoustic scenes that serve the purpose of evaluating room acoustical simulations. At the time of this thesis, this was by far the most comprehensive database serving this purpose, and allows developers and users of room acoustical simulation software packages to evaluate, compare, and improve their algorithms and simulation results on the fly.

Chapter 6 described the first round robin on room acoustical simulation and evaluation, and thus laid the foundation and methodical framework for future across-algorithm comparisons. The results pointed out several ares for future improvement in the chain of room acoustical simulation – from the acquisition of the input data to the implementation of sound field simulations.

Appendix A introduced AKtools – an open toolbox for the acquisition, inspection, and processing of acoustic signals. This modular code framework not only fosters reproducible research by making it easier to detail research methods, but also helps to sustain knowledge at the Audio Communication Group, and makes sophisticated measurement and signal processing techniques available for student projects.

Appendix B detailed the development of the open design and 3D-printable PIRATE ear-plug for conducting individualized binaural recordings. The simple and easy to build design of the ear plugs in combination with free cultural license under which they are provided and measurement techniques from AKtools, considerably decreased the effort that is necessary to conduct reliable individual binaural recordings.

7.2 Future perspectives

7.2.1 Acquisition of individual binaural signals

A measurement protocol for acquiring individual binaural (room) impulse response was detailed in Chapter 2. Due to the time consuming procedure, however, this approach does not appear to be feasible outside the applied research. The question thus arises of how to obtain individualized binaural recordings? Accelerated acoustic measurement techniques based on interleaved exponential sweeps[1] and adaptive filters[2] are a vital option for the anechoic case and reduce the measurement time for full-spherical head-related impulse

[1] J.-G. Richter and J. Fels (**2019**). "On the influence of continuous subject rotation during high-resolution head-related transfer function measurements" *IEEE/ACM Trans. on Audio, Speech, and Language Proc.*

[2] F. Brinkmann, et al. (**2019**). "A cross-evaluated database of measured and simulated HRTFs including 3D head meshes, anthropometric features, and headphone impulse responses" *J. Audio Eng. Soc. (in print).*

response (HRIRs) datasets to a couple of minutes. Although direct acoustic measurements have without a doubt the highest external validity, their limiting factor is that HRIRs will most often only be measured for a fixed head-above torso orientation. This will be problematic in dynamic virtual acoustic simulations, where an unnatural coloration might occur due to the head centric coordinate system that is used in virtual acoustic reality. For example an HRIR measured below the listener will be used for the auralization of a virtual source in front of the listener in case the listener is looking upwards, which would cause an unnatural high frequency damping due to the acoustic shadow of the torso. While listener view activated adaptive filters were proposed to accelerate the measurement of multiple head-above-torso orientations[3], this has so far only been used for a single source position. As an alternative to acoustic measurements, HRIRs for different head-above-torso orientations can also be numerically simulated based on 3D surface meshes[4]. Sufficiently accurate, fast, and mobile scanning methods are available to obtain individual meshes[5], and HRIRS can be simulated quickly by means of cloud or parallel computing[6,7]. However, generating meshes for all possible head-above-torso orientations would require either an unfeasible number of separate scans, or tedious manual mesh post-processing steps. At this point, the previously suggested methodologies might be extended in two ways to obtain full-spherical individual HRIR sets for a multitude of head-above-torso orientations.

View activated multi-channel adaptive filters could be used to acoustically measure HRIRs for all desired head-above-torso orientations on a relatively sparse spherical sampling grid in a first step. A spatially continuous representation could be obtained in a second step by means of HRIR pre-processing for spherical harmonics interpolation[8,9]. This would require a spherical array of about 25 loudspeakers to obtain HRIRs of good quality, but the combination of view activated multichannel adaptive filters and spherical harmonics processing requires additional research. Alternatively HRIRs could be simulated based on head only 3D surface meshes, and the effect of the torso could be taken into account by means of a parametric model. In the simplest case, such a model could consist of an IIR low-pass filter to account for the torso shadowing and an IIR delay to account for the shoulder reflection. While the cut-off frequency of the low-pass filter would be mainly determined by the source elevation, the IIR delay would depend on the head orientation and source position. More sophisticated models might incorporate additional HRIRs depending on the direction of arrival of the shoulder reflection. Noteworthy, the *Dirac 3D Rendering*[10] engine claims to account for head-above-torso orientation, however no details nor any patents were published since the announcement of *dynamic HRTF*[11]. Such accelerated acquisition techniques along with HRIR individualization based on perceptual selection[12], anthropometric features[13], or machine learning[14] might pave the way for future individualized binaural experiences.

[3] J. He, et al. (**2018**). "Fast continuous measurement of HRTFs with unconstrained head movements for 3D audio" *J. Audio Eng. Soc*

[4] F. Brinkmann, et al. (**2017**c). "A high resolution and full-spherical head-related transfer function database for different head-above-torso orientations" *J. Audio Eng. Soc*

[5] M. Dinakaran, et al. (**2018**). "Perceptually motivated analysis of numerically simulated head-related transfer functions generated by various 3D surface scanning systems" in *IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*.

[6] H. Ziegelwanger, et al. (**2015**a). "MESH2HRTF: An open-source software package for the numerical calculation of head-related transfer functions" in *22nd International Congress on Sound and Vibration*.

[7] T. Huttunen and A. Vanne (**2017**). "End-to-end process for HRTF personalization" in *142nd AES Convention, e-Brief 348*.

[8] C. Schörkhuber, et al. (**2018**). "Binaural rendering of Ambisonics signals via magnitude least sqaures" in *Fortschritte der Akustik – DAGA 2018*.

[9] D. L. Alon, et al. (**2018**). "Sparse head-related transfer function representation with spatial aliasing cancellation" in *IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*.

[10] Dirac (**2019**). *Dynamic 3D Rendering* `https://www.dirac.com/3d-audio-rendering` (Assessed Mar. 2019).

[11] Dirac (**2016**). *Dynamic HRTF technology* `https://www.dirac.com/news` (Assessed Mar. 2019).

[12] B. F. G. Katz and G. Parseihian (**2012**). "Perceptually based head-related transfer function database optimization" *J. Acoust. Soc. Am. (Express Letter)*.

[13] S. Ghorbal, et al. (**2017**). "Pinna morphologiacal parameters influence HRTF sets" in *DAFX-17*.

[14] F. Shahid, et al. (**2018**). "AI DevOps for large-scale HRTF prediction and evaluation: an end to end pipeline" in *AES Int. Conf. on Audio for Virtual and Augmented Reality (AVAR)*.

To this point, however, only anechoic binaural signals were considered. Since directly measuring reverberant binaural signals in situ at the listener's ears is not an option for consumer applications, room acoustical simulations could be used to model these signals based on individual(ized) HRIRs. Chapter 6 proved that plausible auralizations can be rendered from room acoustical simulations, but also showed that obvious deficits remain – for example pertaining to the coloration and source position – which might not be acceptable for some applications. In such cases the external validity of acoustic measurements is appealing, and various approaches could be used to obtain individual signals from spatial room impulse responses, i.e., room impulse responses measured with spatially distributed microphone arrays. Using spherical microphone arrays and spherical harmonics processing, BRIRs can be obtained based on a plane wave decomposition of the sound field and a spherical harmonics representation of HRIR datasets[15]. Unfortunately, an impractically high number of microphones would be needed to correctly capture all sound field aspects. Bernschütz[16] concluded that a reduced spherical harmonics order of $N = 11$ (requiring at least $(N+1)^2$ microphones) can render perceptually transparent BRIRs if using optimal radial filters and HRTF subsampling with either a spectral compensation or Bandwidth Extension for Microphone Arrays (BEMA), while an order of $N = 7$ was deemed a good compromise between perceptual quality and technical feasibility. The order requirements to correctly model the binaural receiver might be further relaxed if considering HRIR pre-processing techniques introduced above. The required resolution of the spatial sound field and possible pre-processing techniques to maintain this resolution even at low spherical harmonics orders, however, might be subject to future investigations. Besides spherical harmonics processing, perceptually motivated sound field synthesis techniques such as Directional Audio Coding (DirAC) and the Spatial Decomposition Method (SDM) are also appealing. They aim at estimating the direction of arrival and – in the case of DirAC – diffuseness from four, six, or more microphones[17,18]. Although frequently used, the quality of these techniques deservers more attention and an evaluation against a true binaural reference is required before considering them for the evaluation of room acoustical simulations. The huge benefit of microphone array based BRIR calculation is the decoupling of the spatial sound field and the binaural listener, which means that BRIRs for arbitrary head orientations can be rendered from a single array impulse response.

The approaches discussed so far enable individual binaural synthesis with three degrees of freedom (3DOF), i.e. listener at a fixed position inside a room head rotations along the horizontal plane (yaw), median plane (pitch), and frontal plane (roll). While this appears to be sufficient for evaluating room acoustical simulations, many applications demand 6DOF to simulate a listener that is moving through the virtual environment. Possibilities for realizing such moving listeners are room acoustical simulations in real-time[19] or

[15] B. Rafaely and A. Avni (**2010**). "Interaural cross correlation in a sound field represented by spherical harmonics" *J. Acoust. Soc. Am.*

[16] B. Bernschütz (**2016**). "Microphone arrays and sound field decomposition for dynamic binaural synthesis" Doctoral Thesis.

[17] V. Pulkki, et al., eds. (**2018**). *Parametric time-frequency domain spatial audio*.

[18] S. Tervo, et al. (**2013**). "Spatial decomposition method for room impulse responses" *J. Audio Eng. Soc.*

[19] S. Pelzer, et al. (**2014**). "Integrating real-time room acoustics simulation into a cad modeling software to enhance the architectural design process" *Building Acoustics*.

fading between BRIRS measured at multiple positions[20].

[20] A. Neidhardt and N. Knoop (**2017**). "Binaural walk-through scenarios with actual self-walking using an HTC vive" in *Fortschritte der Akustik – DAGA 2017*.

7.2.2 *Room acoustical simulation*

The round robin detailed in Chapter 6 showed that improvements are not only required in the field of room acoustical simulation algorithms, but also with respect to the acquisition of the required input data – in particular the acoustic description of the surfaces. So far, one-octave or third-octave random incidence absorption and scattering coefficients are used almost exclusively, whereas angle dependent and complex-valued coefficients would be required for a physically correct representation. Since measuring these data *in situ* appears unrealistic in the near future due to the detrimental effect of reflections from other surfaces[21], the question is how this gap should be closed? Because the required data could be measured under laboratory conditions, it would be reasonable to add this data to already existing tables[22]. In addition, an open format for storage and exchange – which is currently missing – could help to raise the acceptance for using such new data, as was the case with the Spatially Oriented Format for Acoustics (SOFA)[23]. Generating an extended database with angle dependent coefficients would without a doubt be beneficial, however, Chapter 6 also showed that using tabulated values comes with an uncertainty that will most likely cause deviations between the simulated and actual acoustics of an environment. Theoretically, material probes could be taken from a room of interest and be measured in the laboratory, but this appears to be infeasible in practice. Since in situ measurements can be conducted in a narrow frequency band, one possibility would be to use this technique to find a best matching material from tabulated vales. In other cases, where detailed material properties are known (e.g. flow resistance, thickness), analytic models could complement (in situ) measurements[24].

[21] E. Brandão, et al. (**2015**). "A review of the in situ impedance and sound absorption measurement techniques" *Acta Acust. united Ac.*

[22] e.g., Physikalisch-Technische Bundesanstalt (PTB) (**2012**). `https://www.ptb.de/cms/fileadmin/internet/fachabteilungen/abteilung_1/1.6_schall/1.63/abstab_wf.zip` (last checked Mar. 2019).

[23] AES Standards Comittee (**2015**). *AES69-2015: AES standard for file exchange - Spatial acoustic data file format*.

Regarding the acoustic modeling algorithms, the most obvious problem is the disregard of wave based effects – namely diffraction and scattering. Although methods for modeling diffraction for image sources[25] and ray tracing[26] are available, only one software package that participated in the round robin made use of such algorithms. Another way to at least partially account for diffraction would be to use room models with a frequency dependent level of detail by automatically discarding objects and structures that are small compared to the wave length of interest[27,28]. Scattering, on the other hand, remains a challenge for algorithms based on geometrical acoustics. While the scattering of rather simple structures like the periodic diffusor used in the round robin (Scene 1, Simple reflection) might be simulated with Svensson's edge diffraction sources mentioned above, more complex and non-regular structures can so far only be approximated by assigning measured[29] of analytical calculated[30] coefficients to a single and structurally simplified surface.

Using wave based simulation algorithms that numerically solve

[24] R. Opdam, et al. (**2015**). "Angle-dependent reflection factor measurements for finite samples with an edge diffraction correction" in *Inter-Noise*.

[25] U. P. Svensson, et al. (**1999**). "An analytic secondary source model of edge diffraction impulse responses" *J. Acoust. Soc. Am.*

[26] A. Pohl and U. M. Stephenson (**2014**). "Combining higher order reflections with diffractions without explosion of computation time: The sound particle radiosity method" in *Proc. of the EAA Joint Symposium on Auralization and Ambisonics*.

[27] S. Siltanen, et al. (**2008**). "Geometry reduction in room acoustics modeling" *Acta Acusts. united Ac.*

[28] S. Drechsler (**2014**). "An algorithm for automatic geometry simplification for room acoustical simulation based on regression planes" *Acta Acust. united Ac.*

[29] ISO 17497-1 (**2004**). *Sound-scattering properties of surfaces. Part 1: Measurement of the random-incidence scattering coefficient in a reverberation room.*

[30] J. J. Embrechts, et al. (**2001**). "Determination of the scattering coefficient of random rough diffusing surfaces for room acoustics applications" *Acta Acust. united Ac.*

the wave equation, diffraction and scattering are inherently modeled. In most cases, the Finite Difference Time Domain (FDTD) method is used where sound waves are propagated through a volumetric mesh with boundary conditions that represent the acoustic environment[31]. Ideally, such algorithms could replace geometrical acoustics modeling, however, due to memory requirements and computational cost their use is currently restricted to the low frequency end, which raises the question of how hybrid wave based and geometrical simulations can be combined[32]. As an alternative, more abstract parametric sound field representations can be generated from band-limited wave based simulations[33] with applications lying rather in the fields of virtual reality and game audio than acoustic planning and environmental noise mapping. Whereas modeling wave effects is inherent to FDTD methods, modeling directional sources and receivers is ongoing research[34,35,36]. Considering recent progress in this area, wave based simulations can be expected to play a more important role in future (room) acoustical simulations, which might decrease the high inter-algorithm variance for low frequencies (cf. Fig 6.5) that was observed in the round robin conducted in Chapter 6.

Apart from the simulation algorithm the user expertise was shown to affect the results already in the second round robin on room acoustical simulation[37] – a fact that also showed up in the round robin detailed in Chapter 6. Two aspects might be considered to guarantee more consistent and valid results across different users: First, concise and easy to understand in-app help/information as well as default parameters that are tuned for high precision rather than a short computation time. Second, auto-detection of critical aspects related to the room models, such as an obstructed direct sound path or the presence of many small surfaces in case an algorithm that does not account for diffraction.

7.2.3 Evaluation and prediction of room acoustic impression

The currently conducted round robin (cf. Chapter 6) evaluated the simulation algorithms with respect to their ability to predict the acoustics of not yet existing venues, which is one important use case. This was done by providing absorption and scattering coefficients mostly based on tabulated values without adjusting them to measured acoustic quantities such as the reverberation time. While this is without a doubt a relevant scenarios for room acoustical simulation, the errors related to uncertainty in the input data can not entirely be separated from those related to the simulation algorithms themselves. To account for this, at least partially, future studies might repeat parts of the round robin (scenes 9 – 11, cf. Fig 5.1) with fitted absorption coefficients, which might better reflect the true potential of the tested software packages.

Another important area of room acoustical simulation is the improvement of existing venues. With respect to this, it is vital to correctly predict the physical and perceptual effects of an acoustic treat-

[31] B. Hamilton (**2016**). "Finite difference and finite volume methods for wave-based modelling of room acoustics" Doctoral Thesis.

[32] A. Southern, et al. (**2013**). "Room impulse response synthesis and validation using a hybrid acoustic model" *IEEE Transactions on Audio, Speech, and Language Processing*.

[33] K. W. Godin, et al. (**2018**). "Wave acoustics in a mixed reality shell" in *AES Int. Conf. on Audio for Virtual and Augmented Reality (AVAR)*.

[34] R. Mehra, et al. (**2014**). "Source and listener directivity for interactive wave-based sound propagation" *Trans. Vis. Comput. Graphics*.

[35] S. Bilbao and B. Hamilton (**2019**). "Directional sources in wave-based acoustic simulation" *IEEE/ACM Trans. Audio, Speech, Language Process.*.

[36] S. Bilbao, et al. (**2019**). "Local time-domain spherical harmonic spatial encoding for wave-based acoustic simulation" *IEEE Signal Processing Letters*

[37] I. Bork (**2000**). "A comparison of room simulation software – The 2nd round robin on room acoustical computer simulation" *Acta Acust. united Ac.*

ment – an aspect, which is currently not fully covered by the data base introduced in Chapter 5 and the evaluation conducted in Chapter 6. The best way to evaluate this would be to gather input data (i.e., a room model, surface properties, and (binaural) room impulse responses) of a room before and after an acoustic treatment. For simulating the room before the acoustical treatment, the software users would be allowed to tune the simulation results to measured room impulse response properties[38,39], whereas for simulating the sound field in the acoustically treated environment no further fitting would be allowed and the absorption and scattering properties of the new materials would have to be taken as specified by the manufacturer or from previously conducted laboratory measurements. In this kind of evaluation the acoustically treated environment had to be compared against the acoustics of the untreated environment, instead of directly comparing simulations and measurements.

Apart from simulating room acoustics, modeling outdoor sound propagation becomes increasingly important for environmental noise mapping and virtual reality applications. In these cases the relevance of diffraction increases because direct sound paths will more often be blocked by buildings or other objects at least for some sound sources inside such scenes. As a consequence, a scene for diffraction around a rectangular cuboid was already added to the database introduced in Chapter 5 after the round robin was conducted.

Because the perceptual evaluation of room acoustical simulations is time consuming it is of high interest to find reliable predictors for the evoked auditory events. This was one goal in the development of room acoustical parameters[40]. They are, however, mostly restricted to an analysis of the temporal energy decay measured with omnidirectional receivers, and their ability to predict binaural and spatial aspects of auditory perception are thus limited. Binaural modeling appears to be a more promising approach as it draws predictions from an analysis of binaural audio streams or impulse responses rather than single channel impulse responses of an omnidirectional receiver[41]. Existing models are already capable of predicting for instance the reverberance, clarity[42], listener envelopment, apparent source width[43], and speech intelligibility[44]. Due to their stream-based nature, most models are also capable of predicting effects of the audio content, a task in which classical room acoustical parameters fail. Nevertheless, not all perceptual aspects can be covered by existing models and additional research is required to predict important perceptions, for instance, related to the tone color or an irregular temporal energy decay[45].

[38] S. Pelzer and M. Vorländer (**2013**). "Inversion of a room acoustics model for the determination of acoustical surface properties in enclosed spaces" in *Proc. Mtgs. Acoust.*

[39] C. L. Christensen and J. H. Rindel (**2014**). "Estimating absorption of materials to match room model against existing room using a genetic algorithm" in *Forum Acusticum*.

[40] ISO 3382-1 (**2009**). *Measurement of room acoustic parameters – Part 1: Performance spaces*.

[41] J. Braasch (**2005**). "Modeling of binaural hearing" in *Communication Acoustics*, edited by J. Blauert.

[42] J. van Dorp Schuitman, et al. (**2013**). "Deriving content-specific measures of room acoustic perception using a binaural, nonlinear auditory model" *J. Acoust. Soc. Am.*

[43] S. Klockgether and S. van de Par (**2014**). "A model for the prediction of room acoustical perception based on the just noticeable differences of spatial perception" *Acta Acust. united Ac.*

[44] O. Kokabi, et al. (**2018**). "Segmentation of binaural room impulse responses for speech intelligibility prediction" *J. Acoust. Soc. Am.*

[45] S. Weinzierl, et al. (**2018**). "A measuring instrument for the auditory perception of rooms: The Room Acoustical Quality Inventory (RAQI)" *J. Acoust. Soc. Am.*

Acknowledgements

First of all, I would like to thank Stefan Weinzierl and Alexander Lindau for introducing me to scientific working through teaching and countless discussions, and for their trust in my work.

I am also grateful to all my co-authors, my colleagues at the Audio Communication Group, and master students for your collaboration and help without which this thesis would not have been possible.

Without any doubt, this also holds for the workshop staff at TU Berlin, RWTH Aachen University, and Carl von Ossiettzky Universtiy Oldenburg, and for everyone who contributed to the countless acoustic measurements that were conducted for this thesis.

My thanks also go to everyone who participated in the listening experiments – especially those of you, who took part in the tests for authenticity. My conclusions build upon your judgement.

I am much obligated to the general public, and the German research foundation in specific for funding my work (DFG WE 4057/3-1, and WE 4057/3-2). You might wanna think about more permanent positions in research.

I also greatly appreciate all contributions and feedback to AKtools, that helped to extend and improve the functionality and documentation of the code.

Last but not least, a big thank you to my family and friends for your support and understanding.

List of publications

This thesis was written cumulatively and contains the following scientific research papers. For brevity, only seven key publications were included in the written version.

Peer reviewed journal publications

Fabian Brinkmann, Lukas Aspöck, David Ackermann, Steffen Lepa, Michael Vorländer, and Stefan Weinzierl (**2019**). "A round robin on room acoustical simulation and auralization." *J. Acoust. Soc. Am.*, **145**(4), 2746–2760. DOI: 10.1121/1.5096178 [featured in Chapter 6]

Fabian Brinkmann, Lukas Aspöck, David Ackermann, Rob Opdam, Michael Vorländer, and Stefan Weinzierl (**2019**). "A benchmark for room acoustical simulation. Concept and database." *Acta Acust. United Ac. (under review)*. [featured in Chapter 5]

Fabian Brinkmann, Manoj Dinakaran, Robert Pelzer, Peter Grosche, Daniel Voss, and Stefan Weinzierl (**2019**). "A cross-evaluated database of measured and simulated HRTFs including 3D head meshes, anthropometric features, and headphone impulse responses" *J. Audio Eng. Soc. (Engineering Report – in print)*.

Omid Kokabi, Fabian Brinkmann, and Stefan Weinzierl (**2019**). "Prediction of speech intelligibility using pseudo-binaural room impulse responses." *J. Acoust. Soc. Am. (Express Letter)*, **145**(4), EL329–EL333. DOI: 10.1121/1.5099169.

Omid Kokabi, Fabian Brinkmann, and Stefan Weinzierl (**2018**). "Segmentation of binaural room impulse responses for speech intelligibility prediction." *J. Acoust. Soc. Am.* **144**(5), 2793–2800. DOI: 10.1121/1.5078598.

David Ackermann, Christoph Böhm, Fabian Brinkmann, and Stefan Weinzierl (**2018**). "The acoustical effect of musicians' movements during musical performances." *Acta Acusts. united Ac.* **105**(2):356–367. DOI: 10.3813/AAA.919319.

Fabian Brinkmann, Alexander Lindau, and Stefan Weinzierl (**2017**). "On the authenticity of individual dynamic binaural synthesis." *J. Acoust. Soc. Am.* **142**(4), 1784–1795. DOI: 10.1121/1.5005606. [featured in Chapter 2]

Fabian Brinkmann, Alexander Lindau, Stefan Weinzierl, Steven van de Par, Markus Müller-Trapet, Rob Opdam, and Michael Vorländer (**2017**). "A high resolution and full-spherical head-related transfer function database for different head-above-torso orientations." *J. Audio Eng. Soc.* **65**(10), 841–848. DOI: 10.17743/jaes.2017.0033. [featured in Chapter 4]

Zamir Ben-Hur, Fabian Brinkmann, Jonathan Sheaffer, Stefan Weinzierl, and Boaz Rafaely (**2017**). "Spectral equalization in binaural signals represented by order-truncated spherical harmonics." *J. Acoust. Soc. Am.* **141**(6), 4087–4096. DOI: 10.1121/1.4983652.

Fabian Brinkmann, Reinhild Roden, Alexander Lindau, and Stefan Weinzierl (**2015**). "Audibility and interpolation of head-above-torso orientation in binaural technology." *IEEE J. Sel. Topics Signal Process.* **9**(1), 931–942. DOI: 10.1109/jstsp.2015.2414905. [featured in Chapter 3]

Alexander Lindau, Vera Erbes, Steffen Lepa, Hans-Joachim Maempel, Fabian Brinkmann, and Stefan Weinzierl (**2014**). "A Spatial Audio Quality Inventory (SAQI)." *Acta. Acust. united Ac.* **100**(5), 984–994. DOI: 10.3813/AAA.918778.

Peer reviewed conference publications

Manoj Dinakaran, Fabian Brinkmann, Stine Harder, Robert Pelzer, Peter Grosche, Rasmus R. Paulsen, and Stefan Weinzierl (**2018**). "Perceptually motivated analysis of numerically simulated head-related transfer functions generated by various 3D surface scanning systems." In *IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)* pp. 551–555. DOI: 10.1109/ICASSP.2018.8461789.

Fabian Brinkmann, Alexander Lindau, Martina Vrhovnik, and Stefan Weinzierl (**2014**). "Assessing the authenticity of individual dynamic binaural synthesis." In *EAA Joint Symposium on Auralization and Ambisonics*, Berlin, Germany, pp. 62–66. DOI: 10.14279/depositonce-11.

Conference publications

Florian Denk, Fabian Brinkmann, Alfred Stirnemann and Briger Kollmeier (**2019**). "The PIRATE – an anthropometric earPlug with exchangeable microphones for Individual Reliable Acquisition of Transfer functions at the Ear canal entrance" *Fortschritte der Akustik – DAGA 2019*, Rostock, Germany, pp. 635–638. DOI: 10.5281/zenodo.2574395. [featured in Appendix B]

Fabian Brinkmann, and Stefan Weinzierl (**2018**). "Comparison of head-related transfer functions pre-processing techniques for spherical harmonics decomposition." In *AES Int. Conf. on Audio for Virtual and Augmented Reality (AVAR)*, Redmond, USA.

Fabian Brinkmann, Vera Erbes, and Stefan Weinzierl (**2018**). "Extending the closed form image source model for source directivity." In *Fortschritte der Akustik – DAGA 2018*, Munich, Germany, pp. 1298–1301.

Silke Bögelein, Fabian Brinkmann, David Ackerman, and Stefan Weinzierl (**2018**). "Localization cues of a spherical head model." In *Fortschritte der Akustik – DAGA 2018*, Munich, Germany, pp. 347–350.

Fabian Brinkmann and Stefan Weinzierl (**2017**). "AKtools – An Open Software Toolbox for Signal Acquisition, Processing, and Inspection in Acoustics." *142nd AES Convention*, Berlin, Germany, e-Brief 309. URL: `http://www.aes.org/e-lib/browse.cfm?elib=18685`. [featured in Appendix A]

Felicitas Fiedler, David Ackerman, Fabian Brinkmann, Martin Schneider, and Stefan Weinzierl (**2017**). "Entwicklung und Evaluation eines Mikrofonarrays für die Aufnahme von räumlichen Schallfeldern nach dem Motion-Tracked Binaural (MTB) Verfahren." In *Fortschritte der Akustik – DAGA 2017*, Kiel, Germany, pp. 1115–1117.

Gunar Schlenstedt, Fabian Brinkmann, Sönke Pelzer, and Stefan Weinzierl (**2016**). "Perzeptive Evaluation transauraler Binauralsynthese unter Berücksichtigung des Wiedergaberaums." In *Fortschritte der Akustik – DAGA 2016*, Aachen, Germany, pp. 561–564.

Manoj Dinakaran, Peter Grosche, Fabian Brinkmann, and Stefan Weinzierl (**2016**). "Extraction of anthropometric measures from 3D-meshes for the individualization of head-related transfer functions." In *140th AES Convention*, Paris, France, Convention Paper 9579.

Fabian Brinkmann, Alexander Lindau, Markus Müller-Trapet, Michael Vorländer, and Stefan Weinzierl (**2015**). "Cross-validation of measured and modeld head-related transfer functions." In *Fortschritte der Akustik – DAGA 2015*, Nürnberg, Germany, pp. 1118–1121.

Braxton Boren, Michele Geronazzo, Fabian Brinkmann, and Edgar Choueiri (**2015**). "Coloration metrics for headphone equalization." In *21st Int. Conf. on Auditory Display.*, Graz, Austria, pp. 29–34.

Mina Fallahi, Fabian Brinkmann, and Stefan Weinzierl (**2015**). "Simulation and analysis of measurement techniques for the fast acquisition of head-related transfer functions." In *Fortschritte der Akustik – DAGA 2015*, Nürnberg, Germany, pp. 1107–1110.

Alexander Fuß, Fabian Brinkmann, Thomas Jürgensohn, and Stefan Weinzierl (**2015**). "Ein vollsphärisches Multikanalmesssystem zur schnellen Erfassung räumlich hochaufgelöster, individueller kopfbezogener Übertragungsfunktionen." In *Fortschritte der Akustik – DAGA 2015*, Nürnberg, Germany, pp. 1114–1117.

Fabian Brinkmann, Reinhild Roden, Alexander Lindau, and Stefan Weinzierl (**2014**). "Audibility of head-above-torso orientation in head-related transfer functions." In *Forum Acusticum*, Kraków, Poland.

Reinhild Roden, Fabian Brinkmann, Alexander Lindau, and Stefan Weinzierl (**2014**). "Resolution and interpolation of different head-above-torso orientations in head-related transfer functions (in German)." In *Fortschritte der Akustik – DAGA 2014*, Oldenburg, Germany, pp. 568–569.

Alexander Lindau, Fabian Brinkmann, and Stefan Weinzierl (**2014**). "Sensory Profiling of Individual and Non-individual Dynamic Binaural Synthesis Using the Spatial Audio Quality Inventory." In *Forum Acusticum*, Kraków, Poland.

Fabian Brinkmann, Alexander Lindau, Stefan Weinzierl, Gunnar Geissler, and Steven van de Par (**2013**). "A high resolution head-related transfer function database including different orientations of head above the torso." In *AIA-DAGA 2013, International Conference on Acoustics*, Merano, Italy, pp. 569–599.

The author of this thesis substantially contributed to the planning, conduction, and writing of all publications that are mentioned above. Moreover, the following contributions to open source, open data, and open design projects are considered to be part of this thesis.

Open source Fabian Brinkmann (**2016**). *AKtools – An Open Software Toolbox for Signal Acquisition, Processing, and Inspection in Acoustics*. Available from `www.ak.tu-berlin.de/AKtools`.

Open source Simon Ciba, Fabian Brinkmann, and Alexander Lindau (**2014**). *WhisPER. A MATLAB toolbox for performing quantitative and qualitative listening tests*. DOI: `10.14279/depositonce-31.3`.

Open data Lukas Aspöck, Fabian Brinkmann, David Ackerman, Stefan Weinzierl, and Michael Vorländer (**2019**). *BRAS – A benchmark for room acoustical simulation.* DOI: `10.14279/depositonce-6726.2`.

Open data Fabian Brinkmann, Manoj Dinakaran, Robert Pelzer, Jan Joschka Wohlgemuth, Fabian Seipel, Daniel Voss, Peter Grosche, and Stefan Weinzierl (**2019**). *The HUTUBS head-related transfer function (HRTF) database.* DOI: `10.14279/depositonce-8487`.

Open data Fabian Brinkmann, Alexander Lindau, Stefan Weinzierl, Gunnar Geissler, Steven van de Par, Markus Müller-Trapet, Rob Opdam, and Michael Vorländer (**2017**). *The FABIAN head-related transfer function data base*. DOI: `10.14279/depositonce-5718.2`

Open design Florian Denk, Fabian Brinkmann, Alfred Stirnemann, and Birger Kollmeier Birger (**2019**). *The PIRATE – an anthropometric earPlug with exchangeable microphones for Individual Reliable Acquisition of Transfer functions at the Ear canal entrance.* (Data set, Version 1), DOI:`10.5281/zenodo.2574395`.

Bibliography

AES Standards Comittee (**2015**). *AES69-2015: AES standard for file exchange - Spatial acoustic data file format* (Audio Engineering Society, Inc.).

Ahnert Feistel Media Group. "EASE – User's Guide & Tutorial (Version 4.3), Acesssed in November 2018" `http://www.afmg-support.de/SoftwareDownloadBase/AFMG/EASE/EASE_4.3_Tutorial_English.pdf` (Accessed: Mar. 2019).

Jens Ahrens, Mark R. P. Thomas, and Ivan J. Tashev (**2012**). "HRTF magnitude modeling using a non-regularized least-squares fit of spherical harmonics coefficients on incomplete data," in *APSIPA Annual Summit and Conference*, Hollywood, USA.

V. Ralph Algazi, Carlos Avendano, and Richard O. Duda (**2001**a). "Elevation localization and head-related transfer function analysis at low frequencies," *J. Acoust. Soc. Am.* **109**(3), 1110–1122, DOI: `10.1121/1.1349185`.

V. Ralph Algazi, Richard O. Duda, Ramani Duraiswami, Nail A. Gumerov, and Zhihui Tang (**2002**). "Approximating the head-related transfer function using simple geometric models of the head and torso," *J. Acoust. Soc. Am.* **112**(5), 2053–2064, DOI: `10.1121/1.1508780`.

V. Ralph Algazi, Richard O. Duda, Dennis M. Thompson, and Carlos Avendano (**2001**b). "The CIPIC HRTF database," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, New York, pp. 99–102, DOI: `10.1109/ASPAA.2001.969552`.

David Lou Alon, Zamir Ben-Hur, Boaz Rafaely, and Ravish Mehra (**2018**). "Sparse head-related transfer function representation with spatial aliasing cancellation," in *IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, Calgary, Canada, pp. 6792–6796.

Lidia Álvarez-Morales, M. Galindo, S. Girón, T. Zamarreño, and R. M. Cibrián (**2016**). "Acoustic characterisation by using different room acoustics software tools: A comparative study," *Acta Acust. united Ac.* **102**, 578–590, DOI: `10.3813/AAA.918975`.

Areti Andreopoulou, Durant R. Begault, and Brian F. G. Katz and (**2015**). "Inter-laboratory round robin HRTF measurement comparison," *IEEE J. Sel. Topics Signal Process.* **9**(5), 895 – 906, DOI: `10.1109/JSTSP.2015.2400417`.

Areti Andreopoulou and Brian F. G. Katz (**2017**). "Identification of perceptually relevant methods of inter-aural time difference estimation," *J. Acoust. Soc. Am.* **142**(2), 588–598, DOI: `10.1121/1.4996457`.

Christiane Antweiler, Aulis Telle, Peter Vary, and Gerald Enzner (**2011**). "Perfect-sweep NLMS for time-variant acoustic system identification," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Prague, Czech Republic, DOI: `10.1109/ICASSP.2012.6287930`.

Lukas Aspöck, Fabian Brinkmann, David Ackerman, Stefan Weinzierl, and Michael Vorländer (**2018**). "GRAS – Ground Truth for Room Acoustical Simulation" https://dx.doi.org/10.14279/depositonce-6726.

Lukas Aspöck, Fabian Brinkmann, David Ackerman, Stefan Weinzierl, and Michael Vorländer (**2019**). "BRAS – A Benchmark for Room Acoustical Simulation" https://dx.doi.org/10.14279/depositonce-6726.2.

Lukas Aspöck, Rob Opdam, and Michael Vorländer (**2016**). "Acquisition of boundary conditions for a room acoustic simulation comparison," in *ISMRA*, Bueanos Aires, Argentina.

Robert Baumgartner, Piotr Majdak, and Bernhard Laback (**2014**). "Modeling sound-source localization in sagittal planes for human listeners," *J. Acoust. Soc. Am.* **136**(2), 791–802, DOI: 10.1121/1.4887447.

Victor Benichoux, Marc Rébillat, and Romain Brette (**2016**). "On the variation of interaural time differences with frequency," *J. Acoust. Soc. Am.* **139**(4), 1810–1821.

Benjamin Bernschütz (**2013**). "A spherical far field HRIR/HRTF compilation of the Neumann KU 100," in *AIA-DAGA 2013, International Conference on Acoustics*, Merano, Italy, pp. 592–595.

Benjamin Bernschütz (**2016**). "Microphone arrays and sound field decomposition for dynamic binaural synthesis," Doctoral Thesis, Technische Universität Berlin, Berlin, Germany, DOI: 10.14279/depositonce-5082.

Marco Berzborn, Ramona Bomhardt, Johannes Klein, Jan-Gerrit Richter, and Michael Vorländer (**2017**). "The ITA-Toolbox: An open source MATLAB toolbox for acoustic measurements and signal processing," in *Fortschritte der Akustik – DAGA 2017*, Kiel, Germany, pp. 222–225.

Stefan Bilbao and Brian Hamilton (**2019**). "Directional sources in wave-based acoustic simulation," *IEEE/ACM Trans. Audio, Speech, Language Process.* **27**(2), 415–428, DOI: 10.1109/TASLP.2018.2881336.

Stefan Bilbao, Archontis Politis, and Brian Hamilton (**2019**). "Local time-domain spherical harmonic spatial encoding for wave-based acoustic simulation," *IEEE Signal Processing Letters* In print, DOI: 10.1109/LSP.2019.2902509.

Jens Blauert (**1997**). *Spatial Hearing. The psychophysics of human sound localization*, revised ed. (MIT Press, Camebridge, Massachusetts).

Ingolf Bork (**2000**). "A comparison of room simulation software – The 2nd round robin on room acoustical computer simulation," *Acta Acust. united Ac.* **86**, 943–956.

Ingolf Bork (**2005**a). "Report on the 3rd round robin on room acoustical computer simulation - Part I: Measurements," *Acta Acust. united Ac.* **91**, 740–752.

Ingolf Bork (**2005**b). "Report on the 3rd round robin on room acoustical computer simulation - Part II: Calculations," *Acta Acust. united Ac.* **91**, 753–763.

Bjarke P. Bovbjerg, Flemming Christensen, Pauli Minnaar, and Xiaoping Chen (**2000**). "Measuring the head-related transfer functions of an artificial head with a high directional resolution," in *109th AES Convention, Preprint*, Los Angeles, USA.

Jonas Braasch (**2005**). "Modeling of binaural hearing," in *Communication Acoustics*, edited by Jens Blauert (Springer, Berlin, Heidelberg), pp. 75–108, DOI: 10.1007/b139075.

Eric Brandão, Arcanjo Lenzi, and Stephan Paul (**2015**). "A review of the in situ impedance and sound absorption measurement techniques," *Acta Acust. united Ac.* **101**(3), 443–463, DOI: 10.3813/AAA.918840.

Albert S. Bregman (**1994**). *Auditory scene analysis. The perceptual organization of sound*, first paperback edition ed. (MIT Press, Camebridge, USA).

Fabian Brinkmann, David Ackerman, and Stefan Weinzierl (**2018**). "The first international Round Robin on room acoustical simulation and auralization. Perceptual evaluation," *J. Acoust. Soc. Am. (Submitted)* .

Fabian Brinkmann, Manoj Dinakaran, Robert Pelzer, Peter Grosche, Daniel Voss, and Stefan Weinzierl (**2019**). "A cross-evaluated database of measured and simulated HRTFs including 3D head meshes, anthropometric features, and headphone impulse responses," *J. Audio Eng. Soc. (in print)* .

Fabian Brinkmann, Alexander Lindau, Markus Müller-Trapet, Michael Vorländer, and Stefan Weinzierl (**2015**a). "Cross-validation of measured and modeld head-related transfer functions," in *Fortschritte der Akustik – DAGA 2015*, Nürnberg, Germany, pp. 1118–1121.

Fabian Brinkmann, Alexander Lindau, Martina Vrhovnik, and Stefan Weinzierl (**2014**a). "Assessing the authenticity of individual dynamic binaural synthesis," in *Proc. of the EAA Joint Symposium on Auralization and Ambisonics*, Berlin, Germany, pp. 62–68, DOI: 10.14279/depositonce-11.

Fabian Brinkmann, Alexander Lindau, and Stefan Weinzierl (**2017**a). "On the authenticity of individual dynamic binaural synthesis," *J. Acoust. Soc. Am.* **142**(4), 1784–1795, DOI: 10.1121/1.5005606.

Fabian Brinkmann, Alexander Lindau, Stefan Weinzierl, Gunnar Geissler, and Steven van de Par (**2013**). "A high resolution head-related transfer function database including different orientations of head above the torso," in *AIA-DAGA 2013, International Conference on Acoustics*, Merano, Italy, pp. 596–599.

Fabian Brinkmann, Alexander Lindau, Stefan Weinzierl, Gunnar Geissler, Steven van de Par, Markus Müller-Trapet, Rob Opdam, and Michael Vorländer (**2017**b). "The FABIAN head-related transfer function data base" https://dx.doi.org/10.14279/depositonce-5718.2.

Fabian Brinkmann, Alexander Lindau, Stefan Weinzierl, Steven van de Par, Markus Müller-Trapet, Rob Opdam, and Michael Vorländer (**2017**c). "A high resolution and full-spherical head-related transfer function database for different head-above-torso orientations," *J. Audio Eng. Soc* **65**(10), 841–848, DOI: 10.17743/jaes.2017.0033.

Fabian Brinkmann, Reinhild Roden, Alexander Lindau, and Stefan Weinzierl (**2014**b). "Audibility of head-above-torso orientation in head-related transfer functions," in *Forum Acusticum*, Europ. Acoust. Assoc., Kraków, Poland.

Fabian Brinkmann, Reinhild Roden, Alexander Lindau, and Stefan Weinzierl (**2015**b). "Audibility and interpolation of head-above-torso orientation in binaural technology," *IEEE J. Sel. Topics Signal Process.* **9**(5), 931–942, DOI: 10.1109/jstsp.2015.2414905.

Fabian Brinkmann and Stefan Weinzierl (**2016**). "AKtools – An open toolbox for acoustic signal acquisition, processing, and inspection" www.ak.tu-berlin.de/AKtools.

Fabian Brinkmann and Stefan Weinzierl (**2017**). "AKtools – An open software toolbox for signal acquisition, processing, and inspection in acoustics," in *142nd AES Convention, Convention e-Brief 309*, Berlin, Germany.

Fabian Brinkmann and Stefan Weinzierl (**2018**). "Comparison of head-related transfer functions preprocessing techniques for spherical harmonics decomposition," in *AES Int. Conf. on Audio for Virtual and Augmented Reality (AVAR)*, Redmond, USA.

Robert Bristow-Johnson (**1994**). "The equivalence of various methods of computing biquad coefficients for audio parametric equalizers," in *97th AES Convention, Preprint*, San Francisco, USA.

Douglas S. Brungart and William M Rabinowitz (**1999**). "Auditory localization of nearby sources. Head-related transfer functions," *J. Acoust. Soc. Am.* **106**(3), 1465–1479, DOI: 10.1121/1.427180.

Douglas S. Brungart, Brian D. Simpson, and Alexander J. Kordik (**2005**). "The detectability of head-tracker latency in virtual audio displays," in *Eleventh Meeting of the International Conference on Auditory Display (ICAD)*, Limmerick, Ireland, pp. 37–42.

Michał Bujacz, Piotr Skulimowski, and Paweł Strumiłło (**2012**). "Naviton - a prototype mobility aid for auditory presentation of three-dimensional scenes to the visually impaired," *J. Audio Eng. Soc.* **60**(9), 696–708.

Densil Cabrera, Jianyang Xun, and Martin Guski (**2016**). "Calculating reverberation time from impulse responses: A comparison of software implementations," *Acoustics Australia* **44**(2), 369–378, DOI: 10.1007/s40857-016-0055-6.

James Carpenter and John Bithell (**2000**). "Bootstrap confidence invervals: when, which, what? A practical guide for medical statisticians," *Statistics in Medicine* **19**(9), 1141–1164, DOI: 10.1002/(SICI)1097-0258(20000515)19:9<1141::AID-SIM479>3.0.CO;2-F.

Patrick Chevret (**2015**). "Advantage of the incoherent uniform theory of diffraction for acoustic calculations in open-plan offices," *J. Acoust. Soc. Am.* **137**(1), 94–104, DOI: 10.1121/1.4904527.

Claus Lynge Christensen and Jens Holger Rindel (**2014**). "Estimating absorption of materials to match room model against existing room using a genetic algorithm," in *Forum Acusticum*, Kraków, Poland.

Simon Ciba, Fabian Brinkmann, and Alexander Lindau (**2014**). "WhisPER. A MATLAB toolbox for performing quantitative and qualitative listening tests" https://dx.doi.org/10.14279/depositonce-31.2.

P. Cignoni, C. Rocchini, and R. Scopigno (**1998**). "Metro: Measuring error on simplified surfaces," *Computer Graphics Forum* **17**(2), 167–174, DOI: 10.1111/1467-8659.00236.

Marcella Cipriano, Arianna Astolfi, and David Pelegrin-Garcia (**2017**). "Combined effect of noise and room acoustics on vocal effort in simulated classrooms," *J. Acoust. Soc. Am.* **141**(1), EL51–EL56, DOI: 10.1121/1.4973849.

R. Ciskowski and C Brebbia (**1991**). *Boundary Element Methods in Acoustics* (Elsevier Applied Science, London, UK).

Florian Denk, Stephan M. A. Ernst, Stephan D. Ewert, and Birger Kollmeier (**2018**). "Adapting hearing devices to the individual ear acoustics: Database and target response correction functions for various device styles," *Trends in Hearing* **22**, 2331216518779313, DOI: 10.1177/2331216518779313.

Florian Denk, Jan Heeren, Stephan D. Ewert, Birger Kollmeier, and Stephan M. A. Ernst (**2017**). "Controlling the head position during individual HRTF measurements and its effect on accuracy," in *Fortschritte der Akustik – DAGA 2017*, Kiel, Germany, pp. 1085–1088.

Joseph G. Desloge (**2014**). "pa-wavplay for 32-bit and 64-bit" https://de.mathworks.com/matlabcentral/fileexchange/47336-pa-wavplay-for-32-bit-and-64-bit.

Pascal Dietrich, Bruno Masiero, and Michael Vorländer (**2013**). "On the optimization of the multiple exponential sweep method," *J. Audio Eng. Soc.* **61**(3), 113–124.

DIN 33402-2 (**2005**). *Ergonomics - Human body dimensions - Part 2: Values* (Beuth, Berlin, Germany).

DIN IEC/TS 60318-7:2011 (**2005**). *Electroacoustics – Simulators of human head and ear – Part 7: Head and torso simulator for the measurement of hearing aids* (Beuth, Berlin, Germany).

Manoj Dinakaran, Fabian Brinkmann, Stine Harder, Robert Pelzer, Peter Grosche, Rasmus R. Paulsen, and Stefan Weinzierl (**2018**). "Perceptually motivated analysis of numerically simulated head-related transfer functions generated by various 3D surface scanning systems," in *IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, Calgary, Canada, pp. 551–555.

Dirac (**2016**). "Dynamic HRTF technology" https://www.dirac.com/news (Assessed Mar. 2019).

Dirac (**2019**). "Dynamic 3D Rendering" https://www.dirac.com/3d-audio-rendering (Assessed Mar. 2019).

Stefan Drechsler (**2014**). "An algorithm for automatic geometry simplification for room acoustical simulation based on regression planes," *Acta Acust. united Ac.* **100**(5), 956–963, DOI: 10.3813/AAA.918775.

Ramani Duraiswami, Dimitry N. Zotkin, and Nail A. Gumerov (**2004**). "Interpolation and range extrapolation of HRTFs," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Montreal, Canada, pp. IV 45–48, DOI: 10.1109/ICASSP.2004.1326759.

Jean Jacques Embrechts, Dominique Archambeau, and Guy-Bart Stan (**2001**). "Determination of the scattering coefficient of random rough diffusing surfaces for room acoustics applications," *Acta Acust. united Ac.* **87**(4), 482–494.

Jean Jacques Embrechts and Alexis Billon (**2011**). "Theoretical determination of the random-incidence scattering coefficients of infinite rigid surfaces with a periodic rectangular roughness profile," *Acta Acust. united Ac.* **97**(4), 607–617, DOI: 10.3813/AAA.918441.

Gerald Enzner, Christiane Antweiler, and Sascha Spors (**2013**). "Acquisition and representation of head-related transfer functions," in *The technology of binaural listening*, edited by Jens Blauert, Modern acoustics and signal processing, 1 ed. (Springer, Heidelberg et al.), pp. 57–92, DOI: 10.1007/978-3-642-37762-4.

Vera Erbes, Frank Schultz, Alexander Lindau, and Stefan Weinzierl (**2012**). "An extraaural headphone system for optimized binaural reproduction," in *Fortschritte der Akustik – DAGA 2012*, Darmstadt, Germany, pp. 313–314.

Michael J. Evans, James A. S. Angus, and Anthony I. Tew (**1998**). "Analyzing head-related transfer function measurements using surface spherical harmonics," *J. Acoust. Soc. Am.* **104**(4), 2400–2411, DOI: 10.1121/1.423749.

Mina Fallahi, Fabian Brinkmann, and Stefan Weinzierl (**2015**). "Simulation and analysis of measurement techniques for the fast acquisition of head-related transfer functions," in *Fortschritte der Akustik – DAGA 2015*, Nürnberg, Germany, pp. 1107–1110.

Matthias Frank (**2013**). "Phantom sources using multiple loudspeakers in the horizontal plane," Doctoral Thesis, University of Music and Performing Arts, Graz, Austria.

Alexander Fuß, Fabian Brinkmann, Thomas Jürgensohn, and Stefan Weinzierl (**2015**). "Ein vollsphärisches multikanalmesssystem zur schnellen erfassung räumlich hochaufgelöster, individueller kopfbezogener übertragungsfunktionen," in *Fortschritte der Akustik – DAGA 2015*, pp. 1114–1117.

Mark Bill Gardner (**1973**). "Some monaural and binaural facets of median plane localization," *J. Acoust. Soc. Am.* **54**(6), 1489–1495, DOI: 10.1121/1.1914447.

Walter Gautschi (**2012**). *Numerical analysis*, 2nd edition ed. (Birkhäuser, Basel, Switzerland).

Matthias Geier, Jens Ahrens, and Sascha Spors (**2008**). "The sound scape renderer: A unified spatial audio reproduction framework for arbitrary rendering methods," in *124th AES Convention, Preprint 7330*, Amsterdam, The Netherlands.

Klaus Genuit (**1984**). "Ein Modell zur Beschreibung von Außenohrübertragungseigenschaften," Doctoral Thesis, Technische Hochschule, Aachen, Germany.

Slim Ghorbal, Théo Auclair, Catherine Soladié, and Renaud Séguier (**2017**). "Pinna morphologiacal parameters influence HRTF sets," in *DAFX-17*, Edinburgh, UK.

Keith W. Godin, Ryan Rohrer, John Snyder, and Nikunj Raghuvanish (**2018**). "Wave acoustics in a mixed reality shell," in *AES Int. Conf. on Audio for Virtual and Augmented Reality (AVAR)*, Redmond, USA.

M. Guldenschuh, Alois Sontacchi, and Franz Zotter (**2008**). "HRTF modelling in due consideration variable torso reflections," in *Acoustics*, Paris, France.

Martin Guski and Michael Vorländer (**2014**). "Comparison of noise compensation methods for room acoustic impulse response evaluations," *Acta Acustica united with Acustica* **100**, 320–327.

Brian Hamilton (**2016**). "Finite difference and finite volume methods for wave-based modelling of room acoustics," Doctoral Thesis, University of Edinburgh, Edinburgh, UK.

Dorte Hammershøi and Henrik Møller (**2005**). "Binaural technique – basic methods for recording, synthesis, and reproduction," in *Communication Acoustics*, edited by Jens Blauert, Springer, Berlin, Heidelberg, pp. 223–254.

Klaus Hartung, Jonas Braasch, and Susanne J Sterbing (**1999**). "Comparison of different methods for the interpolation of head-related transfer functions," in *16th Int. AES Conference*, Rovaniemi, Finland, pp. 319–329.

Jianjun He, Rishabh Ranjan, Woon-Seng Gan, Nitesh Kumar Chaudhary, Nguyen Duy Hai, and Rishabh Gupta (**2018**). "Fast continuous measurement of HRTFs with unconstrained head movements for 3D audio," *J. Audio Eng. Soc* **66**(11), 884–900, DOI: 10.17743/jaes.2018.0050.

Etienne Hendrickx, Peter Stitt, Jean-Christophe Messonnier, Jean-Marc Lyzwa, Brian FG Katz, and Catherine de Boishéraud (**2017**). "Influence of head tracking on the externalization of speech stimuli for non-individualized binaural synthesis," *J. Acoust. Soc. Am.* **141**(3), 2011–2023, DOI: 10.1121/1.4978612.

Timo Hiekkanen, Aki Mäkivirta, and Matti Karjalainen (**2009**). "Virtualized listening tests for loudspeakers," *J. Audio Eng. Soc.* **57**(4), 237–251.

Kunikazu Hirosawa, Kazuhiro Takashima, Hiroshi Nakagawa, Makoto Kon, Aki Yamamoto, and Walter Lauriks (**2009**). "Comparison of three measurement techniques for the normal absorption coefficient of sound absorbing materials in the free field," *J. Acoust. Soc. Am.* **126**(6), 3020–3027, DOI: 10.1121/1.3242355.

Maarten Hornikx (**2016**). "Ten questions concerning computational urban acoustics," *Building and Environment* **106**, 409 – 421, DOI: 10.1016/j.buildenv.2016.06.028.

Maarten Hornikx, Manfred Kaltenbacher, and Steffen Marburg (**2015**). "A platform for benchmark cases in computational acoustics," *Acta Acust. united Ac.* **101**(4), 811–820.

Joop J. Hox (**2010**). Quantitative Methodology *Multilevel Analysis. Techniques and Apllications*, 2 ed. (Routledge, New York, Hove).

Robert Humphrey (**2008**). "Playrec – Multichannel Matlab audio" http://www.playrec.co.uk.

Tomi Huttunen and Antti Vanne (**2017**). "End-to-end process for HRTF personalization," in *142nd AES Convention, e-Brief 348*, Berlin, Germany.

Tomi Huttunen, Antti Vanne, Stine Harder, Rasmus Reinhold Paulsen, Sam King, Lee Perry-Smith, and Leo Kärkkäinen (**2014**). "Rapid generation of personalized HRTFs," in *55th Int. AES. Conf.: Spatial Audio*, Helsinki, Finland.

IEC 60268-4 (**2014**). *Sound system equipment – Part 4: Microphones*.

ISO 10534-2 (**1998**). *Determination of sound absorption coefficient and impedance in impedance tubes – Part 2: Transfer-function method* (International Organization for Standards, Geneva, Switzerland).

ISO 17497-1 (**2004**). *Sound-scattering properties of surfaces. Part 1: Measurement of the random-incidence scattering coefficient in a reverberation room* (International Organization for Standards, Geneva, Switzerland).

ISO 17497-2 (**2012**). *Sound-scattering properties of surfaces. Part 2: Measurement of the directional diffusion coefficient in a free field* (International Organization for Standards, Geneva, Switzerland).

ISO 3382-1 (**2009**). *Measurement of room acoustic parameters – Part 1: Performance spaces* (International Organization for Standards, Geneva, Switzerland).

ISO 354 (**2003**). *Measurement of sound absorption in a reverberation room* (International Organization for Standards, Geneva, Switzerland).

ITU-R BS.1116-3 (**2015**). *Methods for the subjective assessment of small impairments in audio systems* (ITU, Geneva, Switzerland).

ITU-R BS.1534-3 (**2015**). *Methods for the subjective assessment of intermediate quality level of audio systems* (ITU, Geneva, Switzerland).

Cheol-Ho Jeong, Doheon Lee, Sébastien Santurette, and Jeong-Guon Ih (**2014**). "Influence of impedance phase angle on sound pressures and reverberation times in a rectangular room," *J. Acoust. Soc. Am.* **135**(2), 712–723, DOI: 10.1121/1.4861839.

Craig Jin, Pierre Guillon, Nicolas Eapain, Reza Zolfaghari, André van Schaik, Anthony I. Tew, Carl Hetherington, and Jonathan Thorpe (**2014**). "Creating the sidney york morphological and acoustic recordings of ears database," *IEEE Trans. on Multimedia* **16**(1), 37–46, DOI: 10.1109/tmm.2013.2282134.

Zora Schärer Kalkandjiev and Stefan Weinzierl (**2015**). "The influence of room acoustics on solo music performances: An empirical investigation," *Psychomusicology: Music, Mind, and Brain* **25**(3), 195–207, DOI: 10.1037/pmu0000065.

Brian F. G. Katz (**2001**). "Boundary element method calculation of individual head-related transfer function. II. impedance effects and comparisons to real measurements," *J. Acoust. Soc. Am.* **110**(5), 2449–2455, DOI: 10.1121/1.1412441.

Brian F. G. Katz (**2004**). "International round robin on room acoustical impulse response analysis software," *Acoustic research letters online* **5**(4), 158–164.

Brian F. G. Katz and Gaëtan Parseihian (**2012**). "Perceptually based head-related transfer function database optimization," *J. Acoust. Soc. Am. (Express Letter)* **131**(2), EL99–EL105, DOI: 10.1121/1.3672641.

D. B. Keele Jr. (**1974**). "Low-frequency loudspeaker assessment by nearfield sound-pressure measurement," *J. Audio Eng. Soc.* **22**(3), 154–162.

Chungeun Kim, Russel Mason, and Tim Brookes (**2013**). "Head movements made by listeners in experimental and real-life listening activities," *J. Audio Eng. Soc.* **61**(6), 425–438.

Ole Kirkeby, Eira T. Seppälä, Asta Kärkkäinen, Leo Kärkkäinen, and Tomi Huttunen (**2007**). "Some effects of the torso on head-related transfer functions," in *122nd AES Convention, Convention Paper 7030*, Vienna, Austria.

Mendel Kleiner, Bengt-Inge Dalenbäck, and Peter Svensson (**1993**). "Auralization – An overview," *J. Audio Eng. Soc.* **41**(11), 861–875.

Stefan Klockgether and Steven van de Par (**2014**). "A model for the prediction of room acoustical perception based on the just noticeable differences of spatial perception," *Acta Acust. united Ac.* **100**(8), 964–971, DOI: 10.3813/AAA.918776.

Tobias Knüttel, Ingo B. Witew, and Michael Vorländer (**2013**). "Influence of "omnidirectional" loudspeaker directivity on measured room impulse responses," *J. Acoust. Soc. Am.* **134**(5), 3654–3662, DOI: 10.1121/1.4824334.

Omid Kokabi, Fabian Brinkmann, and Stefan Weinzierl (**2018**). "Segmentation of binaural room impulse responses for speech intelligibility prediction," *J. Acoust. Soc. Am.* **144**(5), 2793–2800, DOI: 10.1121/1.5078598.

George F. Kuhn (**1977**). "Model for the interaural time differences in the azimuthal plane," *J. Acoust. Soc. Am.* **62**(1), 157–167, DOI: 10.1121/1.381498.

Heinrich Kuttruff (**1991**). "Digital simulation of concert hall acoustics and its applications," *Acoustic Bulletin* **16**(5), 5–8.

Heinrich Kuttruff (**2009**). *Room acoustics*, 5th edition ed. (Spon Press, Oxford, UK).

Timo I. Laakso, Vesa Välimäki, Matti Karjalainen, and Unto K. Laine (**1996**). "Splitting the unit delay," *IEEE Signal Processing Magazine* **13**(1), 30–60, DOI: 10.1109/79.482137.

Erno H. A. Langendijk and Adelbert W. Bronkhorst (**2000**). "Fidelity of three-dimensional-sound reproduction using a virtual auditory display," *J. Acoust. Soc. Am.* **107**(1), 528–537, DOI: 10.1121/1.428321.

Timothy W. Leishman, Sarah Rollins, and Heather M. Smith (**2006**). "An experimental evaluation of regular polyhedron loudspeakers as omnidirectional sources of sound," *J. Acoust. Soc. Am.* **120**(3), 1411–1422, DOI: 10.1121/1.2221552.

Les Leventhal (**1986**). "Type 1 and type 2 errors in the statistical analysis of listening tests," *J. Audio Eng. Soc.* **34**(6), 437–453.

Jerad Lewis and Brian Moss (**2013**). "MEMS microphones, the future of hearing aids," *Analog Dialogue* **47**(11), 1–3.

Alexander Lindau (**2014**). "Binaural resynthesis of acoustical environments. technology and perceptual evaluation," Ph.D. Thesis, Technical Universtiy Berlin, Berlin, Germany, DOI: 10.14279/depositonce-4085.

Alexander Lindau and Fabian Brinkmann (**2012**). "Perceptual evaluation of headphone compensation in binaural synthesis based on non-individual recordings," *J. Audio Eng. Soc.* **60**(1/2), 54–62.

Alexander Lindau, Fabian Brinkmann, and Stefan Weinzierl (**2014**a). "Sensory profiling of individual and non-individual dynamic binaural synthesis using the spatial audio quality inventory," in *Forum Acusticum*, Europ. Acoust. Assoc., Kraków, Poland.

Alexander Lindau, Vera Erbes, Steffen Lepa, Hans-Joachim Maempel, Fabian Brinkmann, and Stefan Weinzierl (**2014**b). "A Spatial Audio Quality Inventory (SAQI)," *Acta Acust. united Ac.* **100**(5), 984–994, DOI: 10.3813/AAA.918778.

Alexander Lindau, Jorgos Estrella, and Stefan Weinzierl (**2010**). "Individualization of dynamic binaural synthesis by real time manipulation of the ITD," in *128th AES Convention, Convention Paper*, London, UK.

Alexander Lindau, Torben Hohn, and Stefan Weinzierl (**2007**). "Binaural resynthesis for comparative studies of acoustical environments," in *122th AES Convention, Convention Paper 7032*, Vienna, Austria.

Alexander Lindau, Linda Kosanke, and Stefan Weinzierl (**2012**). "Perceptual evaluation of model- and signal-based predictors of the mixing time in binaural room impulse responses," *J. Audio Eng. Soc.* **60**(11), 887–898.

Alexander Lindau and Stefan Weinzierl (**2009**). "On the spatial resolution of virtual acoustic environments for head movements on horizontal, vertical and lateral direction," in *EAA Symposium on Auralization*, Espoo, Finland.

Alexander Lindau and Stefan Weinzierl (**2012**). "Assessing the plausibility of virtual acoustic environments," *Acta Acust. united Ac.* **98**(5), 804–810, DOI: 10.3813/AAA.918562.

Shuai Lu, Xiang Yan, Junjie Li, and Weiguo Xu (**2016**). "The influence of shape design on the acoustic performance of concert halls from the viewpoint of acoustic potential of shapes," *Acta Acusts. united Ac.* **102**(6), 1027–1044, DOI: 10.3813/AAA.919017.

Paul Luizard, Jean-Dominique Polack, and Brian F. G. Katz (**2014**). "Sound energy decay in coupled spaces using a parametric analytical solution of a diffusion equation," *J. Acoust. Soc. Am.* **135**(5), 2765–2776, DOI: 10.1121/1.4870706.

Piotr Majdak et al. (**2017**). "Collection of public head-related impulse response data bases" http://sofacoustics.org/data/database/ (checked July 2017).

Francesco Martellotta (**2013**). "Optimizing stepwise rotation of dodecahedron sound source to improve the accuracy of room acoustic measures," *J. Acoust. Soc. Am.* **134**(3), 2037–2048, DOI: 10.1121/1.4817879.

Ricardo San Martín, Ingo B. Witew, M. Arana, and Michael Vorländer (**2007**). "Influence of the source orientation on the measurement of acoustic parameters," *Acta Acust. united Ac.* **93**(3), 387–397.

Bruno Masiero (**2012**). "Individualized binaural technology. measurement, equalization and perceptual evaluation," Doctoral Thesis, RWTH Aachen, Aachen, Germany.

Ken I. McAnally and Russell L. Martin (**2014**). "Sound localization with head movement: implications for 3-d audio displays," *Frontiers in Neuroscience* **8**, 210, DOI: 10.3389/fnins.2014.00210.

Ravish Mehra, Lakulish Antani, Sujeong Kom, and Dinesh Manocha (**2014**). "Source and listener directivity for interactive wave-based sound propagation," *Trans. Vis. Comput. Graphics* **20**(4), 495–503, DOI: 10.1109/TVCG.2014.38.

A. W. Mills (**1958**). "On the minimum audible angle," *J. Acoust. Soc. Am.* **30**(4), 237–246, DOI: 10.1121/1.1909553.

Pauli Minnaar, Krarup Soren Olesen, Flemming Christensen, and Henrik Møller (**2001**). "The importance of head movements for binaural room synthesis," in *International Conference on Auditory Display*, Espoo, Finland, pp. 21–25.

Pauli Minnaar, Jan Plogsties, and Flemming Christensen (**2005**). "Directional resolution of head-related transfer functions required in binaural synthesis," *J. Audio Eng. Soc.* **53**(10), 919–929.

Henrik Møller (**1992**). "Fundamentals of binaural technology," *Appl. Acoust.* **36**, 171–218.

Henrik Møller, Dorte Hammershøi, Clemen Boje Jensen, and Michael Friis Sørensen (**1995**a). "Transfer characteristics of headphones measured on human ears," *J. Audio Eng. Soc.* **43**(4), 203–217.

Henrik Møller, Michael Friis Sørensen, Dorte Hammershøi, and Clemens Boje Jensen (**1995**b). "Head-related transfer functions of human subjects," *J. Audio Eng. Soc.* **43**(5), 300–321.

Henrik Møller, Michael Friis Sørensen, Clemens Boje Jensen, and Dorte Hammershøi (**1996**). "Binaural technique: Do we need individual recordings?," *J. Audio Eng. Soc.* **44**(6), 451–469.

Eckard Mommertz (**1995**). "Angle-dependent in-situ measurements of reflection coefficients using a subtraction technique," *Applied Acoustics* **46**(3), 251 – 263, DOI: 10.1016/0003-682X(95)00027-7.

Alastair H. Moore, Anthony I. Tew, and Rozenn Nicol (**2007**). "Headphone transparification: A novel method for investigating the externalisation of binaural sounds," in *123rd AES Convention, Convention Paper 7166*, New York, USA.

Alastair H. Moore, Anthony I. Tew, and Rozenn Nicol (**2010**). "An initial validation of individualised crosstalk cancellation filters for binaural perceptual experiments," *J. Audio Eng. Soc.* **58**(1/2), 36–45.

Brian C. J. Moore, Simon R. Oldfield, and Gary J. Dooley (**1989**). "Detection and discrimination of spectral peaks and notches at 1 and 8 kHz," *J. Acoust. Soc. Am.* **85**(2), 820–836, DOI: 10.1121/1.397554.

Clifford T. Morgan, A. Chapanis, Jesse S. Cook, and M. W. Lund (**1963**). *Human Engineering Guide to Equipment Design* (McGraw-Hill, New York, USA).

Michael Möser (**2009**). *Engineering acoustics: An Introduction to noise control*, 2 ed. (Springer, Berlin Heidelberg).

Jennifer E. Mossop and John F. Culling (**1998**). "Lateralization of large interaural delays," *J. Acoust. Soc. Am.* **104**(3), 1574–1579, DOI: 10.1121/1.424369.

Swen Müller and Paulo Massarani (**2001**). "Transfer function measurement with sweeps. directors cut including previously unreleased material and some corrections," *J. Audio Eng. Soc.* (Original release) **49**(6), 443–471.

Markus Müller-Trapet, Pascal Dietrich, Marc Aretz, Jan van Gemmeren, and Michael Vorländer (**2013**). "On the in situ impedance measurement with pu-probes — Simulation of the measurement setup," *J. Acoust. Soc. Am.* **134**(2), 1082–1089, DOI: 10.1121/1.4812250.

Kevin R. Murphy and Brett Myors (**1999**). "Testing the hypothesis that treatments have negligible effects: Minimum-effect tests in the general linear model," *J. of Applied Psychology* **84**(2), 234–248.

Shinichi Nakagawa and Holger Schielzeth (**2013**). "A general and simple method for obtaining R^2 from generalized linear mixed-effects models," *Methods in Ecology and Evolution* **4**, 133–142, DOI: 10.1111/j.2041-210x.2012.00261.x.

Annika Neidhardt and Niklas Knoop (**2017**). "Binaural walk-through scenarios with actual self-walking using an HTC vive," in *Fortschritte der Akustik – DAGA 2017*, Kiel, Germany, pp. 283–286.

Rozenn Nicol (**2010**). AES Monograph *Binaural Technology* (Audio Eng. Soc., New York, USA).

Christian Nocke (**2000**). "In-situ acoustic impedance measurement using a free-field transfer function method," *Applied Acoustics* **59**(3), 253 – 264, DOI: 10.1016/S0003-682X(99)00004-3.

Scott G. Norcross, Martin Bouchard, and Gilbert A. Soulodre (**2006**). "Inverse filtering design using a minimal phase target function from regularization," in *121th AES Convention, Convention Paper 6929*, San Francisco, USA.

Antonín Novák, Laurent Simon, František Kadlec, and Pierrick Lotton (**2010**). "Nonlinear system identification using exponential swept-sine signal," *IEEE Trans. on Instrumentation and Measurement* **59**(8), 2220–2229.

Josefa Oberem, Bruno Masiero, and Janina Fels (**2016**). "Experiments on authenticity and plausibility of binaural reproduction via headphones employing different recording methods," *Appl. Acoust.* **114**, 71–78, DOI: 10.1016/j.apacoust.2016.07.009.

Odeon A/S (**2018**). "ODEON Room Acoustics Software User's Manual (2016. Accessed: Mar. 2019)" https://odeon.dk/wp-content/uploads/2017/09/ODEON_Manual.pdf.

Rob Opdam, Diemer de Vries, and Michael Vorländer (**2014**). "Locally or non-locally reacting boundaries: Does it make a significant acoustic difference?," *Building Acoustics* **21**(2), 117–124, DOI: 10.1260/1351-010X.21.2.117.

Rob Opdam, Diemer de Vries, and Michael Vorländer (**2015**). "Angle-dependent reflection factor measurements for finite samples with an edge diffraction correction," in *Inter-Noise*, San Francisco, USA.

Alan V. Oppenheim and Ronald W. Schafer (**2010**). *Discrete-time signal processing*, third edition ed. (Pearson, Upper Saddle et al.).

Alan V. Oppenheim, Ronald W. Schafer, and John R. Buck (**1999**). *Discrete-time signal processing*, 2nd edition ed. (Prentice Hall, Upper Saddle, USA).

Toru Otsuru, Tetsuya Sakuma, and Shinichi Sakamoto (**2005**). "Constructing a database of computational methods for environmental acoustics," *Acoust. Sci. & Tech.* **26**(2), 221–224, DOI: 10.1250/ast.26.221.

Jörn Otten (**2001**). "Factors influencing acoustical localization," Doctoral Thesis, Universität Oldenburg, Oldenburg, Germany.

Stefanie Otto and Stefan Weinzierl (**2009**). "Comparative simulations of adaptive psychometric procedures," in *NAG/DAGA 2009, International Conference on Acoustics*, Rotterdam, Netherland, pp. 1276–1279.

Mathieu Paquier and Vincent Koehl (**2015**). "Discriminability of the placement of supra-aural and circumaural headphones," *Appl. Acoust.* **93**, 130–139, DOI: 10.1016/j.apacoust.2015.01.023.

Antonio Pedrero, Alexander Díaz-Chyla, César Díaz, Sönke Pelzer, and Michael Vorländer (**2014**). "Virtual restoration of the sound of the hispanic rite," in *Forum Acusticum*, Kraków, Poland.

Sönke Pelzer, Marc Aretz, and Michael Vorländer (**2011**). "Quality assessment of room acoustic simulation tools by comparing binaural measurements and simulations in an optimized test scenario (a)," *Acta Acust. united Ac.* **97**(S1), 102–103.

Sönke Pelzer, Lukas Aspöck, Dirk Schröder, and Michael Vorländer (**2014**). "Integrating real-time room acoustics simulation into a cad modeling software to enhance the architectural design process," *Building Acoustics* **4**(2), 113–138, DOI: 10.3390/buildings4020113.

Sönke Pelzer and Michael Vorländer (**2013**). "Inversion of a room acoustics model for the determination of acoustical surface properties in enclosed spaces," in *Proc. Mtgs. Acoust.*, Vol. 19, p. 015115, DOI: 10.1121/1.4800297.

Physikalisch-Technische Bundesanstalt (PTB) (**2012**). https://www.ptb.de/cms/fileadmin/internet/fachabteilungen/abteilung_1/1.6_schall/1.63/abstab_wf.zip (last checked Mar. 2019).

Chris Pike, Frank Melchior, and Tony Tew (**2014**). "Assessing the plausibility of non-individualised dynamic binaural synthesis in a small room," in *AES 55th International Conference*, Helsinki, Finland.

Alexander Pohl and Uwe M. Stephenson (**2014**). "Combining higher order reflections with diffractions without explosion of computation time: The sound particle radiosity method," in *Proc. of the EAA Joint Symposium on Auralization and Ambisonics*, Berlin, Germany, pp. 119–125, DOI: 10.14279/depositonce-20.

Martin Pollow, Khoa-Van Nguyen, Olivier Warusfel, Thibaut Carpentier, Markus Müller-Trapet, Michael Vorländer, and Markus Noisternig (**2012**). "Calculation of head-related transfer functions for arbitrary field points using spherical harmonics decomposition," *Acta Acust. united Ac.* **98**, 72–82, DOI: 10.3813/AAA.918493.

Barteld N. J. Postma and Brian F. G. Katz (**2016**a). "Correction method for averaging slowly time-variant room impulse response measurements," *J. Acoust. Soc. Am.* **140**(1), EL38–EL43, DOI: 10.1121/1.4955006.

Barteld N. J. Postma and Brian F. G. Katz (**2016**b). "Perceptive and objective evaluation of calibrated room acoustic simulation auralizations," *J. Acoust. Soc. Am.* **140**(6), 4326–4337, DOI: 10.1121/1.4971422.

Danièle Pralong and Simon Carlile (**1996**). "The role of individualized headphone calibration for the generation of high fidelity virtual auditory space," *J. Acoust. Soc. Am.* **100**(6), 3785–3793, DOI: 10.1121/1.417337.

Miller S. Puckette et al. (**1997**). "Pure data.," in *Int. Computer Music Conf. (ICMC)*, Thessaloniki, Greece.

Ville Pulkki, Symeon Delikaris-Manias, and Archontis Politis, eds. (**2018**). *Parametric time-frequency domain spatial audio*, 1 ed. (Wiley, Hoboken, USA).

Boaz Rafaely (**2015**). *Fundamentals of spherical array processing*, 1st ed. (Springer, Berlin, Heidelberg, Germany), Erratum: `http://www.ee.bgu.ac.il/~br/fundamentals_of_spherical_array_processing_erratum.pdf`.

Boaz Rafaely and Amir Avni (**2010**). "Interaural cross correlation in a sound field represented by spherical harmonics," *J. Acoust. Soc. Am.* **127**(2), 823–828, DOI: `10.1121/1.3278605`.

Jan-Gerrit Richter and Janina Fels (**2019**). "On the influence of continuous subject rotation during high-resolution head-related transfer function measurements," *IEEE/ACM Trans. on Audio, Speech, and Language Proc.* **27**(4), 730–741, DOI: `10.1109/TASLP.2019.2894329`.

Klaus A. J. Riederer (**2004**). "Part Va: Effect of head movements on measured head-related transfer functions," in *18th Intern. Congress on Acoustics*, Kyoto, Japan, pp. 795–798.

Jens Holger Rindel (**2011**a). "Room acoustic modelling techniques: A comparison of a scale model and a computer model for a new opera theatre," *Building Acoustics* **18**(3-4), 259–280, DOI: `10.1260/1351-010X.18.3-4.259`.

Jens Holger Rindel (**2011**b). "The ERATO project and its contribution to our understanding of the acoustics of ancient theatres," in *The Acoustics of Ancient Theatres Conference*, Patras, Greece.

Scott M. Robeson (**1997**). "Spherical methods for spatial interpolation: Review and evaluation," *Cartography and Geographic Information Systems* **24**(1), 3–20, DOI: `10.1559/152304097782438746`.

Reinhild Roden, Fabian Brinkmann, Alexander Lindau, and Stefan Weinzierl (**2014**). "Auflösung und Interpolation von unterschiedlichen Kopf-über-Torso-Orientierungen in kopfbezogenen Übertragungsfunktionen," in *Fortschritte der Akustik – DAGA 2014*, Oldenburg, Germany, pp. 568–569.

Conor Ryan and Dermot Furlong (**1995**). "Effects of headphone placement on headphone equalisation for binaural reproduction," in *98th AES Convention*, Paris, France.

Tetsuya Sakuma, U. Peter Svensson, Andreas Franck, and Shinichi Sakamoto (**2002**). "A round-robin test programm on wave-based computational methods for room-acoustic analysis," in *Forum Acusticum*, Sevialla, Spain.

Lauri Savioja and U. Peter Svensson (**2015**). "Overview of geometrical room acoustic modeling techniques," *J. Acoust. Soc. Am.* **138**(2), 708–730, DOI: `10.1121/1.4926438`.

Zora Schärer and Alexander Lindau (**2009**). "Evaluation of equalization methods for binaural signals," in *126th AES Convention, Convention Paper*, Munich, Germany.

Carl Schissler, Aaron Nicholls, and Ravish Mehra (**2016**). "Efficient hrtf-based spatial audio for area and volumetric sources," *IEEE Transactions on Visualization and Computer Graphics* **22**(4), 1356–1366, DOI: `10.1109/TVCG.2016.2518134`.

Christian Schörkhuber, Markus Zaunschirm, and Robert Höldrich (**2018**). "Binaural rendering of Ambisonics signals via magnitude least sqaures," in *Fortschritte der Akustik – DAGA 2018*, Munich, Germany, pp. 339–342.

Dirk Schröder (**2011**). "Physically based real-time auralization of interactive virtual environments," Doctoral Thesis, RWTH Aachen, Aachen, Germany.

Dirk Schröder, Michael Vorländer, and U. Peter Svensson (**2010**). "Open acoustic measurements for validating edge diffraction simulation methods," in *Baltic-Nordic Acoustic Meeting (BNAM)*, Bergen, Norway.

See Supplementary materials at `http://dx.doi.org/10.14279/depositonce-8510`.

Katsuaki Sekiguchi and Toshiki Hanyu (**1998**). "Study on acoustic index variations due to small changes in the observation point," in *15th International Congress on Acoustics*, Seattle, USA, pp. 2121–2122.

Faiyadh Shahid, Nikhil Javeri, Kapil Jain, and Shruti Badhwar (**2018**). "AI DevOps for large-scale HRTF prediction and evaluation: an end to end pipeline," in *AES Int. Conf. on Audio for Virtual and Augmented Reality (AVAR)*, Redmond, USA.

Louena Shtrepi, Arianna Astolfi, Sönke Pelzer, Renzo Vitale, and Monika Rychtáriková (**2015**). "Objective and perceptual assessment of the scattered sound field in a simulated concert hall," *J. Acoust. Soc. of Am.* **138**(3), 1485–1497, DOI: `10.1121/1.4929743`.

Samuel Siltanen, Tapio Lokki, Lauri Savioja, and Claus Lynge Christensen (**2008**). "Geometry reduction in room acoustics modeling," *Acta Acusts. united Ac.* **94**(3), 410–418, DOI: `10.3813/AAA.918049`.

Joseph Sinker and Jamie Angus (**2015**). "Efficient compact representation of head related transfer functions for portable game audio," in *AES 56th International Conference: Audio for Games*.

Malcom Slaney (**1998**). "Auditory toolbox. version 2," Technical Report #1998-010 .

Peter L. Søndergaard and Piotr Majdak (**2013**). "The auditory modeling toolbox," in *The technology of binaural listening*, edited by Jens Blauert, Modern acoustics and signal processing, 1 ed. (Springer, Heidelberg et al.), pp. 33–56, DOI: `10.1007/978-3-642-37762-4`.

Alex Southern, Samuel Siltanen, Damian T. Murphy, and Lauri Savioja (**2013**). "Room impulse response synthesis and validation using a hybrid acoustic model," *IEEE Transactions on Audio, Speech, and Language Processing* **21**(9), 1940–1952, DOI: `10.1109/TASL.2013.2263139`.

J. S. Suh and Philip A. Nelson (**1999**). "Measurement of transient response of rooms and comparison with geometrical acoustic models," *J. Acoust. Soc. Am.* **105**(4), 2304–2317.

U. Peter Svensson, Roger I. Fred, and John Vanderkooy (**1999**). "An analytic secondary source model of edge diffraction impulse responses," *J. Acoust. Soc. Am.* **106**(5), 2331–2344, DOI: `10.1121/1.428071`.

Hironori Takemoto, Parham Mokhtari, Hiroaki Kato, Ryouichi Nishimura, and Kazuhiro Iida (**2012**). "Mechanism for generating peaks and notches of head-related transfer functions in the median plane," *J. Acoust. Soc. Am.* **132**(6), 3832–3841, DOI: `10.1121/1.4765083`.

Technical University Berlin (**2016**). "Deposit Once – repository for research data and publications" `https://depositonce.tu-berlin.de`.

Sakari Tervo, Jukka Pätynen, Antti Kuusinen, and Tapio Lokki (**2013**). "Spatial decomposition method for room impulse responses," *J. Audio Eng. Soc.* **61**(1/2), 17–28.

Willard R. Thurlow, John W. Mangels, and Pulip S Runge (**1967**). "Head movements during sound localization," *J. Acoust. Soc. Am.* **42**(2), 489–493, DOI: `10.1121/1.1910605`.

Emiel Tijs (**2013**). "Study and development of an in situ acoustic absorption measurement method," Doctoral Thesis, University of Twente, Enschede, The Netherlands.

Julio Cesar B. Torres, Lukas Aspöck, and Michael Vorländer (**2018**). "Comparative study of two geometrical acoustic simulation models," *J. Brazilian Soc. of Mechanical Sciences and Engineering* **40**(6), 300, DOI: 10.1007/s40430-018-1226-1.

Rendell R. Torres, U. Peter Svensson, and Mendel Kleiner (**2001**). "Computation of edge diffraction for more accurate room acoustics auralization," *J. Acoust. Soc. Am.* **109**(2), 600–610, DOI: 10.1121/1.1340647.

Bernhard Treutwein (**1995**). "Adaptive psychophysical procedures," *Vision Research* **35**(17), 2503–2522, DOI: 10.1016/0042-6989(95)00016-X.

Nicolas Tsingos, I. Carlbom, G. Elko, R. Kubli, and T. Funkhouser (**2002**). "Validating acoustical simulations in the Bell Labs Box," *IEEE Computer Graphics and Applications* **22**(4), 28–37.

Julia Turku, Miikka Vilermo, Eira Seppälä, Monika Pölönen, Ole Kirkeby, Asta Kärkäinnen, and Leo Kärkkäinen (**2008**). "Perceptual evaluation of numerically simulated head-related transfer functions," in *124th AES Convention, Preprint 7489*, Amsterdam, The Netherlands.

Joseph G. Tylka, Rahulram Sridhar, and Edgar Y. Choueiri (**2015**). "A database of loudspeaker polar radiation measurements," in *139th AES Convention, e-Brief 230*, New York, USA.

Vesa Välimäki, Julian Parker, Lauri Savioja, Julius O. Smith, and Jonathan Abel (**2016**). "More than 50 years of artificial reverberation," in *60th In. AES Conf. DREAMS (Dereverberation and Reverberation of Audio, Music, and Speech)*, Leuven, Belgium.

Vesea Välimäki, Julian D. Parker, Lauri Savioja, Julius O. Smith, and Jonathan S. Abel (**2012**). "Fifty years of artificial reverberation," *IEEE Transactions on Audio, Speech, and Language Processing* **20**(5), 1421–1448, DOI: 10.1109/TASL.2012.2189567.

Jasper van Dorp Schuitman, Diemer de Vries, and Alexander Lindau (**2013**). "Deriving content-specific measures of room acoustic perception using a binaural, nonlinear auditory model," *J. Acoust. Soc. Am.* **133**(3), 1572–1585, DOI: 10.1121/1.4789357.

Patrick Vandewalle, Jelena Kovačević, and Martin Vetterli (**2009**). "Reproducible research in signal processing – what, why, and how," *IEEE Signal Processing Magazine* **26**(3), 37–46, DOI: 10.1109/MSP.2009.932122.

Alejandro Osses Vecchi, Armin Kohlrausch, Winfried Lachenmayr, and Eckard Mommertz (**2017**). "Predicting the perceived reverberation in different room acoustic environments using a binaural auditory model," *J. Acoust. Soc. Am.* **141**(4), EL381–EL387, DOI: 10.1121/1.4979853.

Michael Vorländer (**1995**). "International round robin on room acoustical computer simulations," in *15th International Congress on Acoustics*, Trondheim, Norway, pp. 689–692.

Michael Vorländer (**2008**). *Auralization. Fundamentals of acoustics, modelling, simulation, algorithms and acoustic virtual reality*, 1st ed. (Springer, Berlin, Heidelberg, Germany).

Michael Vorländer (**2013**). "Computer simulations in room acoustics: Concepts and uncertainties," *J. Acoust. Soc. Am.* **133**(3), 1203–1213, DOI: 10.1121/1.4788978.

Stefan Weinzierl (**2017**). "An open repository for research data in acoustics (OPERA)," in *Fortschritte der Akustik – DAGA 2017*, Kiel, Germany.

Stefan Weinzierl, Steffen Lepa, and David Ackermann (**2018**). "A measuring instrument for the auditory perception of rooms: The Room Acoustical Quality Inventory (RAQI)," *J. Acoust. Soc. Am.* **144**(3), 1245–1257, DOI: 10.1121/1.5051453.

Stefan Weinzierl, Paolo Sanvito, and Clemens Büttner (**2015**). "The acoustics of Renaissance theatres in Italy," *Acta Acust. united Ac.* **101**(3), 632–641, DOI: 10.3813/AAA.918858.

Stefan Weinzierl and Michael Vorländer (**2015**). "Room acoustical parameters as predictors of room acoustical impression: What do we know and what do we want to know?," *Acoustics Australia* **43**(1), 41–48, DOI: 10.3813/AAA.918858.

Torben Wendt, Steven van de Par, and Stephan D. Ewert (**2014**). "A computationally-efficient and perceptually-plausible algorithm for binaural room impulse response simulation," *J. Audio Eng. Soc* **62**(11), 748 – 766, DOI: 10.17743/jaes.2014.0042.

Hagen Wierstorf (**2014**). "Perceptual assessment of sound field synthesis," Doctoral Thesis, Technische Universität Berlin, Berlin, Germany.

Hagen Wierstorf, Matthias Geier, Alexander Raake, and Sascha Spors (**2011**). "A free database of head-related impulse resonse measurements in the horizontal plane with multiple distances," in *130th AES Convention, Enginieering Brief*.

Frederic L. Wightman and Doris J. Kistler (**1989**a). "Headphone simulation of free field listening. I: Stimulus synthesis," *J. Acoust. Soc. Am.* **85**(2), 858–867, DOI: 10.1121/1.397557.

Frederic L. Wightman and Doris J. Kistler (**1989**b). "Headphone simulation of free field listening. II: Psychological validation," *J. Acoust. Soc. Am.* **85**(2), 868–878, DOI: 10.1121/1.397558.

Frederic L. Wightman and Doris J. Kistler (**1992**). "The dominant role of low-frequency interaural time differences in sound localization," *J. Acoust. Soc. Am.* **91**(3), 1648–1661, DOI: 10.1121/1.402445.

Ingo B. Witew, Pascal Dietrich, Diemer de Vries, and Michael Vorländer (**2010**). "Uncertainty of room acoustics measurements - How many measurement positions are necessary to describe the conditions in auditoria?," in *Proc. Int. Symp. on Room Acoustics (ISRA)*, Melbourne, Australia.

Ning Xiang, Yun Jing, and Alexander C. Bockman (**2009**). "Investigation of acoustically coupled enclosures using a diffusion-equation model," *J. Acoust. Soc. Am.* **126**(3), 1187–1198.

Wen Zhang, Mengqui Zhang, Rodney A. Kennedy, and Thushara D. Abhayapala (**2012**). "On high-resolution head-related transfer function measurements: An efficient sampling scheme," *IEEE Transactions on Audio, Speech and Language Processing* **20**(2), 575–584, DOI: 10.1109/TASL.2011.2162404.

Harald Ziegelwanger, Wolfgang Kreuzer, and Piotr Majdak (**2015**a). "MESH2HRTF: An open-source software package for the numerical calculation of head-related transfer functions," in *22nd International Congress on Sound and Vibration*, Florence, Italy.

Harald Ziegelwanger, Piotr Majdak, and Wolfgang Kreuzer (**2015**b). "Numerical calculation of listener-specific head-related transfer functions and sound localization: Microphone model and mesh discretization," *J. Acoust. Soc. Am.* **138**(1), 208–222, DOI: 10.1121/1.4922518.

Kamila Żychaluk and David H. Foster (**2009**). "Model-free estimation of the psychometric function," *Attention, Perception, & Psychophysics* **71**(6), 1414–1425, DOI: 10.3758/APP.71.6.1414.

Appendices

A
AKtools – an open software toolbox for signal acquisition, processing, and inspection in acoustics

Fabian Brinkmann and Stefan Weinzierl (2017), *142nd AES Convention*, Berlin, Germany, e-Brief 309.
(Accepted manuscript. ©Audio Engineering Societey)

The acquisition, processing, and inspection of audio data plays a central role in the everyday practice of acousticians. However, these steps are commonly distributed among different and often closed software packages making it difficult to document this work. AKtools includes Matlab methods for audio playback and recording, as well as a versatile plotting tool for inspection of single/multi channel data acquired on spherical, and arbitrary spatial sampling grids. Functional blocks cover test signal generation (e.g. pulses, noise, and sweeps), spectral deconvolution, transfer function inversion using frequency dependent regularization, spherical harmonics transform and interpolation among others. Well documented demo scripts show the exemplary use of the main parts, with more detailed information in the description of each method. To foster reproducible research, AKtools is available under the open software European Union Public Licence (EUPL) allowing everyone to use, change, and redistribute it for any purpose: `www.ak.tu-berlin.de/aktools`.

A.1 Introduction

Reproducible research aims at providing all data that are necessary for completely repeating a study. Besides the research data itself – such as audio recordings, impulse responses, or ratings from listening tests – this especially includes the source code for acquiring and processing the data, and generating the graphics[1]. However, publishing all data is particularly challenging in empirical science where a large amount of data is often acquired, followed by a dedicated pre- and post-processing. This also implies that strategies for long

[1] P. Vandewalle, et al. (2009). "Reproducible research in signal processing – what, why, and how" *IEEE Signal Processing Magazine*.

term data availability (data safety) and traceability of changes (versioning) have to be considered in addition to mere publishing. In this context, AKtools are intended to foster reproducible research in fields that rely on the acquisition, processing, and inspection of acoustic signals. Possible strategies for providing the research data itself are given for example by DepositOnce[2] or the Open Repository for Research Data in Acoustics (OPERA - under construction)[3].

[2] Technical University Berlin (**2016**). *Deposit Once – repository for research data and publications* `https://depositonce.tu-berlin.de`.

[3] S. Weinzierl (**2017**). "An open repository for research data in acoustics (OPERA)" in *Fortschritte der Akustik – DAGA 2017*.

A.2 AKtools

A.2.1 Structure

AKtools is structured in three blocks that include (i) demo scripts with examples for functionality and function calls, (ii) the functions themselves, and (iii) demo data for illustrating certain applications. For ease of use, the installation of AKtools, as well as the functional scope and literature references are described in an accompanying read me file. Additionally, all function headers include a detailed list of input and output parameters, and examples for function calls or a reference to a corresponding demo script. To ease the integration of AKtools into existing projects, the functionality is subdivided in small blocks that work on time signals in almost all cases. The time signals are organized in matrices of size $[N\ M\ C]$, where N is the number of samples, M the number of measurements, and C the number of channels. For efficiently switching between the time and frequency domain, AKtools contains functions for the conversion between single and both sided spectra in the case an even and uneven number of frequency bins, or samples, respectively.

A.2.2 Functional Scope

AKtools contains methods for the playback and recording, processing and inspection of acoustic signals, of which some are being introduced in the following. For a comprehensive overview, please refer to the demo scripts that are contained in AKtools, and which include use cases for the vast majority of the functional scope.

The sample precise single and multi channel audio playback and recording is realized by `AKio.m` which uses playrec[4], and pa-wavplay[5]. In combination with the sweep synthesis in the time or frequency domain (`AKsweepFD.m`, `AKseepTD.m`), and the spectral deconvolution (`AKdeconv.m`), this furthermore enables versatile impulse response measurements, e.g. for the identification of non-linear distortion products using exponential sweeps[6], for the adaptive estimation of impulse responses using perfect sweeps[7], or measuring with constant signal-to-noise ratio across frequency using colored sweeps[8]. An example for measuring an impulse response and calibrating the frequency response of the measurement chain is given in `AKmeasureDemo.m`.

[4] R. Humphrey (**2008**). *Playrec – Multichannel Matlab audio* `http://www.playrec.co.uk`.

[5] J. G. Desloge (**2014**). *pa-wavplay for 32-bit and 64-bit* `https://de.mathworks.com/matlabcentral/fileexchange/47336-pa-wavplay-for-32-bit-and-64-bit`.

[6] A. Novák, et al. (**2010**). "Nonlinear system identification using exponential swept-sine signal" *IEEE Trans. on Instrumentation and Measurement*.

[7] C. Antweiler, et al. (**2011**). "Perfect-sweep NLMS for time-variant acoustic system identification" in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*.

[8] S. Müller and P. Massarani (**2001**). "Transfer function measurement with sweeps. directors cut including previously unreleased material and some corrections" *J. Audio Eng. Soc.* (Original release).

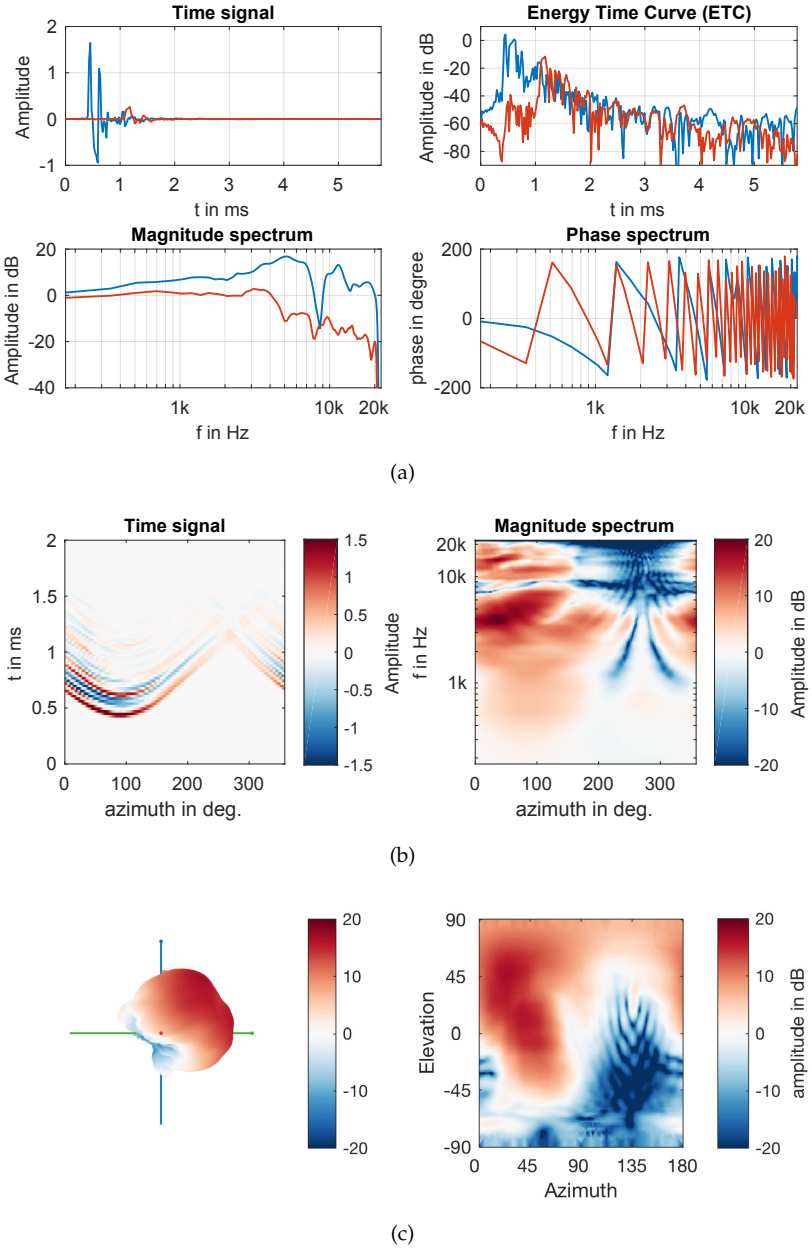
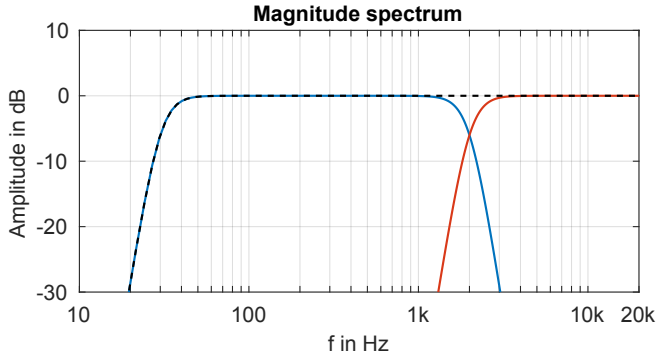


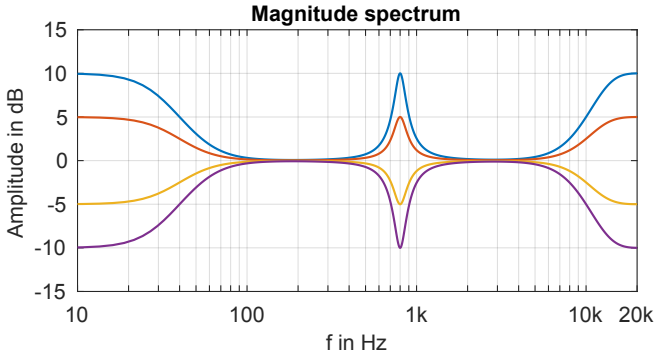
Figure A.1: Examples for graphics generated with AKp.m using head-related impulse responses (HRIRs) and transfer functions (HRTF) from the FABIAN data base. (a) Left ear (blue) and right ear (red) HRIRs and HRTFs of a sound source to the left of the listener. (b) Left ear HRIRs and HRTFs magnitude in the horizontal plane, where 0° azimuth refers to a source in front of the listener, and 90° to a source to the left. (c) Spherical plots of the HRTFs magnitude at 6 kHz, where 90° elevation refers to a source above, and -90° to a source below the listener.

Besides the playback and recording of audio, the quick inspection and verification of signals by means of plotting plays a key roll. The function `AKp.m` holds a multitude of possibilities for this purpose, including simple two-dimensional time and frequency domain plots (Fig. A.1a), and three-dimensional plots of data slices (Fig. A.1b) or spherical data sets (Fig. A.1c). Note that the Matlab code for generating all figures in this manuscript can be found in the Code Appendix.

A simple yet often needed post-processing of acoustic signals is filtering. On this account, `AKfilter.m` provides numerous possibilities including high and low passes, band passes and rejections, crossover networks, and octave filters, as well as high and low shelves,



(a)



(b)

Figure A.2: Examples for filters generated with AKfilter.m: (a) Eighth order Linkwitz-Riley cross-over network (blue, red, $f_c = \{40, 2000\}$ Hz) and the addition of the channels (black). (b) Second order low-shelves and high-shelves ($f_c = \{40, 10.000\}$ Hz, Gain $\{-10, -5, 5, 10\}$ dB), and parametric equalizers with gains of $\{-10, -5, 5, 10\}$ dB ($f_c = 800$ Hz, $Q = 4$).

and parametric equalizers. Some examples are shown in Fig. A.2 – a more comprehensive overview is given in `AKfilterDemo.m`.

Another frequently required task is the transfer function inversion, be it for the equalization of headphones for binaural sound reproduction or the compensation of loudspeakers in room acoustics. AKtools holds the function `AKregulatedInversion.m` which implements a frequency dependent regularized LMS inversion that approaches a minimum phase or linear phase target function. By this means, the precision of the inversion can be restricted for certain frequency ranges, for example to avoid excessive gains in the inverse transfer function[9]. Fig. A.3 shows the signals that are involved in the regularized inversion of a headphone transfer function (HpTF). In this case, the exact compensation of narrow but potentially inaudible notches at approximately 9 kHz and 13 kHz was avoided by means of a regularization function composed from two parametric equalizers.

AKtools moreover includes basic tools for spherical harmonics processing including the DSHT (discrete spherical harmonics transform) that decomposes spherical data sets (e.g. head-related transfer functions – HRTFs) into a weighted superposition of orthogonal spherical basis function shown in Fig. A.4(a). Besides a significant reduction in data this comes with the advantage of a spatially continuous data representation. The DSHT and its inverse according

[9] S. G. Norcross, et al. (**2006**). "Inverse filtering design using a minimal phase target function from regularization" in *121th AES Convention, Convention Paper 6929*.

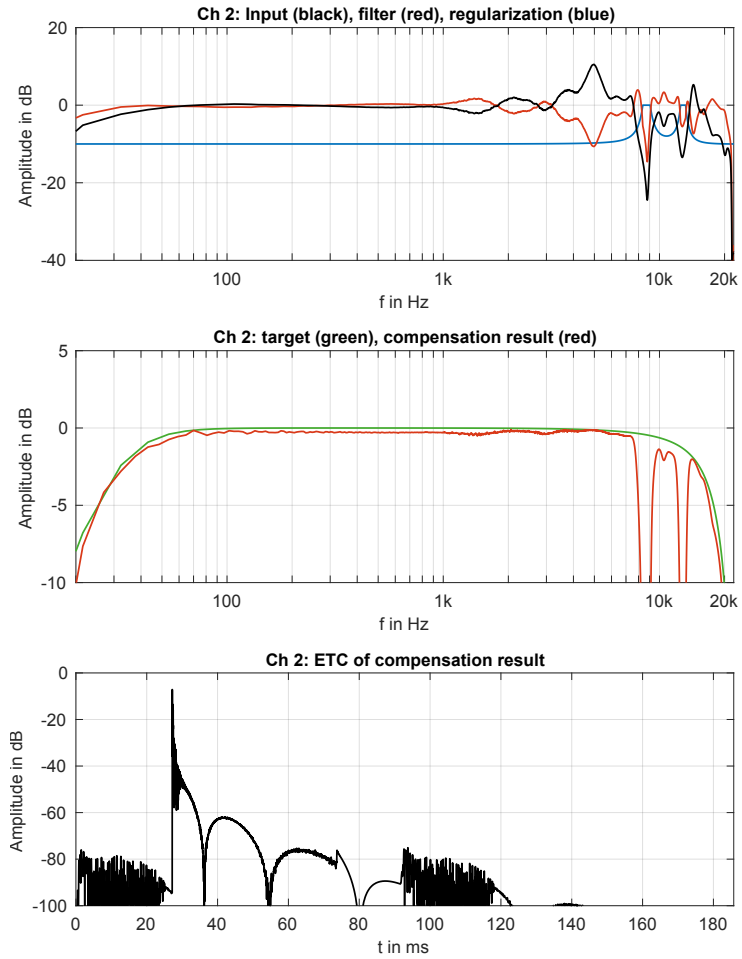


Figure A.3: Regulated inversion on the example of a headphone transfer function (HpTF). Top: HpTF (black), inverted HpTF (red) and regularization function (blue). Middle: target function (green) and inversion result (i.e. convolution of HpTF and inverted HpTF, red). Bottom: Energy time curve ($x[n]^2$) approaching a delayed minimum phase system.

to eq. (3.34) in Rafaely (2015)[10] are implemented in `AKsht.m`, and `AKisht.m` – they can be applied either to complex spectra or separately on magnitude and unwrapped phase values. Examples for an HRTF before and after spherical harmonics transform are depicted in Fig. A.4(b) for different truncation orders. The DSHT requires full spherical data sets, a prerequisite that is often violated by acoustically measured data. A possible solution to this problem is implemented in `AKsphericalCapExtrapolationDemo.m`. Following Ahrens *et al.* (2012)[11], a low order DSHT ($3 \lesssim N \lesssim 4$) is performed on the incomplete data to complete the spatial sampling grid, and a high order DSHT is applied to the completed data in a second step.

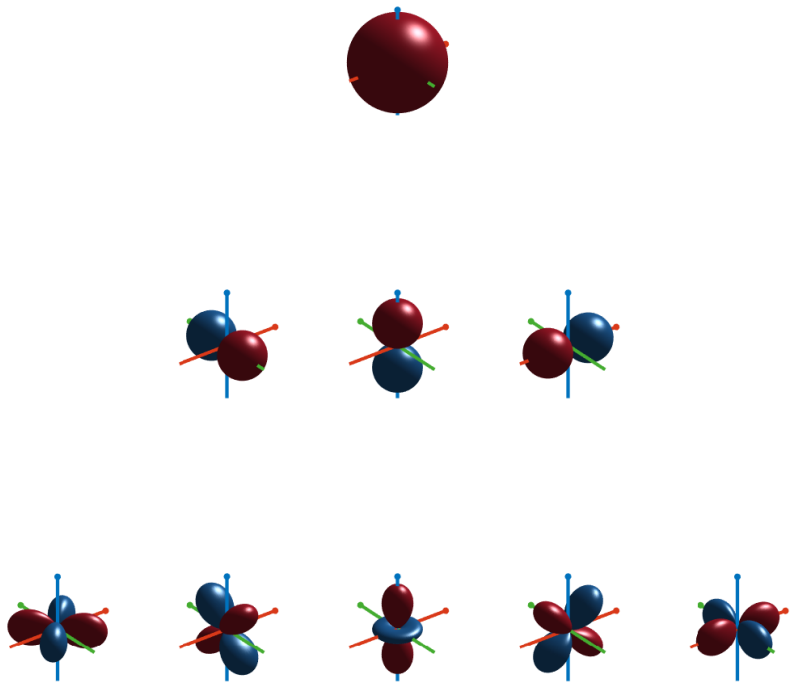
[10] B. Rafaely (**2015**). *Fundamentals of spherical array processing*.

[11] J. Ahrens, et al. (**2012**). "HRTF magnitude modeling using a non-regularized least-squares fit of spherical harmonics coefficients on incomplete data" in *APSIPA Annual Summit and Conference*.

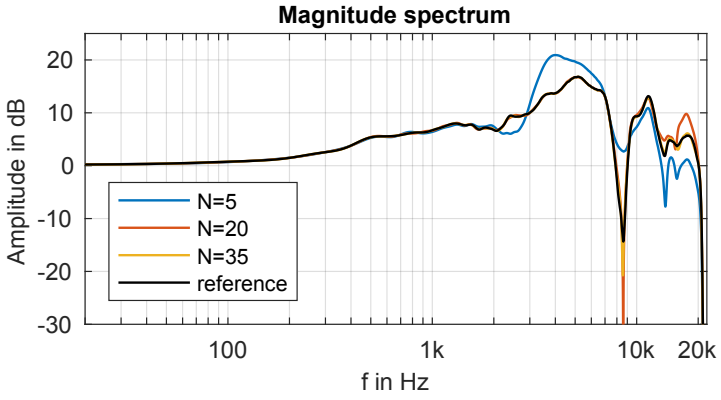
A.2.3 Head-related transfer functions

AKtools are accompanied by the publicly available FABIAN head-related transfer functions data base[12] that holds acoustically measured and numerically simulated HRTFs for eleven head-above torso orientations, and a full-spherical, high resolution sampling grid. For convenience, HRTFs are also provided by means of high order (N=35) spherical harmonics coefficients. HRTFs interpolated to arbitrary source positions and head-above-torso orientations can be obtained using `AKhrirInterpolation.m` which interpolates the source posi-

[12] F. Brinkmann, et al. (**2017**b). *The FABIAN head-related transfer function data base* `https://dx.doi.org/10.14279/depositonce-5718.2`.



(a)



(b)

Figure A.4: (a) Spherical harmonics basis functions for orders 0 (top row), 1 (middle row) and 2 (bottom row). For ease of display, the two leftmost columns show the imaginary part of the complex functions and the 3 rightmost rows the real part. The colors denote positive (red), and negative (blue) values. Red, blue, and green lines point in positive x, y, and z direction, respectively. (b) Head-related transfer function for the left ear of FABIAN and a source to the left before (black), and after discrete spherical harmonics transform with different truncation orders N.

tions in the spherical harmonics domain and the head-above-torso orientation separately on the magnitude and unwrapped phase spectra following Brinkmann *et al.* (2015)[13].

[13] F. Brinkmann, et al. (**2015**b). "Audibility and interpolation of head-above-torso orientation in binaural technology" *IEEE J. Sel. Topics Signal Process.*

A.2.4 Availability

The AKtools are available under the European Union Public Licence (EUPL) granting unrestricted access, and can be obtained from `www.ak.tu-berlin/aktools`. Versioning is of the source code is done by subversion (SVN), and the backup system for data safety is maintained by the Technical University Berlin. AKtools need Matlab 2013 or later including the Signal Processing, and Statistics and Machine Learning Toolbox. However, the majority of functionality is also available without any additional toolboxes. Some demo scripts within

AKtools make use of the FABIAN head-related transfer function data base[14] which is available from `http://dx.doi.org/10.14279/depositonce-5718`.

[14] F. Brinkmann, et al. (2017b). *The FABIAN head-related transfer function data base* `https://dx.doi.org/10.14279/depositonce-5718.2`.

A.3 Summary

The free Matlab toolbox AKtools was introduced which can be used for common tasks in acoustics such as audio playback and recording, impulse response measurements, filtering, transfer function inversion, or HRTF interpolation. AKtools is an ongoing project, that will be extended in the future and welcomes third party contributions.

Code

This appendix demonstrates some AKtools basics and generates all figures that are shown in this manuscript. To start using AKtools download it from `www.ak.tu-berlin.de/aktools` navigate to the AKtools folder in Matlab and run `AKtoolsStart.m`. This source code was tested with AKtools revision 33, and Matlab 2016a.

```
%% ------------------------------------ PLOTTING
% get HRIRs on an evenly sampled grid
g      = AKgreatCircleGrid(-90:2:90, 2, 90);
[l, r] = AKhrirInterpolation(g(:,1), g(:,2), 0, 'measured_ir');

%% get and plot HRIR to the left of the listener
id = AKsubGrid(g, 'any', [90 0]);

AKf(18,9)
AKp([l(:,id) r(:,id)])
print('-dpdf', 'figure1a')

%% get and plot the horizontal plane
id = AKsubGrid(g, 'transverse', 0);

AKf(18,6)
subplot(1,2,1)
    AKp(l(:,id), 't3d', 'y', g(id,1), 'dr', [-1.5 1.5], ...
        'x', [0 2])
    xlabel 'azimuth in deg.'
subplot(1,2,2)
    AKp(l(:,id), 'm3d', 'y', g(id,1), 'dr', [-20 20])
    xlabel 'azimuth in deg.'

print('-dpdf', 'figure1b')

%% plot spherical HRTF dataset at 6 kHz
AKf(18,6)
subplot(1,2,1)
    AKp(l, 's2', 'az', g(:,1), 'el', g(:,2), 'dr', [-20 ...
        20], 'sph_f', 6000, 'labeling', 'off', 'hp_view', ...
        [90 0])
subplot(1,2,2)
    AKp(l, 's5', 'az', g(:,1), 'el', g(:,2), 'dr', [-20 ...
        20], 'sph_f', 6000)
    set(gca, 'xTick', 1:90:361)

print('-djpeg', '-r300', 'figure1c')
```

```
%% ------------------------------------ FILTERING
% genrate dirac signals for filtering
x = AKdirac(2^14, 1);

%% apply high-pass, and cross-over
y = AKfilter(x, 'hp', 30,    0, 44100, 8, 'LR');
y = AKfilter(y, 'xover', 2000, 0, 44100, 8, 'LR');

AKf(12,6)
AKp(squeeze(y), 'm2d', 'x', [10 20000])
AKp(sum(y,3),   'm2d', 'x', [10 20000], 'dr', [10 -30], ...
    'dash', 1)
print('-dpdf', 'figure2a')


%% apply low/high-shelves, and parametric EQs
y = AKfilter(x, 'ls',  40,    [10 5 -5 -10], 44100, 2, 'mid');
y = AKfilter(y, 'hs',  10000, [10 5 -5 -10], 44100, 2, 'mid');
y = AKfilter(y, 'peq', 800,   [10 5 -5 -10], 44100, 4, ...
    'hpl', 'cos');

AKf(12,6)
AKp(y, 'm2d', 'x', [10 20000], 'dr', [-15 15])
print('-dpdf', 'figure2b')
```

```
%% ------------ SPHERICAL HARMONICS BASIS FUNCTIONS
% get spherical harmonics up to order n
N           = 2;
[Ynm, n, m] = AKsh(N, [], g(:,1), 90-g(:,2));

AKf(10)
% plot imaginary part for m<0, and real part for m>=0
for k = 1:numel(n)
    if m(k)<0;
        SH = imag(Ynm(:,k))';
    else
        SH = real(Ynm(:,k))';
    end

    subtightplot(N+1,2*N+1, n(k)*(2*N+1) + N+1 + m(k), 0)
    AKp(SH, 'x7', 'az', g(:,1), 'el', g(:,2), 'cb', 0, ...
        'sph_proc', 'tri')
    title('')
end

print('-dpng', '-r300', 'figure4a')
```

```
%% ------------------- SPERICAL HARMONICS TRANSFORM
% spherical harmonics tranform
% (left ear HRIRs, SH order N=35)
f_nm = AKsht(l, true, [g(:,1) 90-g(:,2)], 35, 'complex');

%% inverse transform, differernt truncation orders
% (source on hprizontal plane, 90 deg. to the left)
N = [5 20 35];
for nn = 1:numel(N);
    l_sh(:,nn) = AKisht(f_nm(1:(N(nn)+1)^2,:), true, [90 ...
        90], 'complex');
end

%% plot reference and order truncated HRTFs
id = AKsubGrid(g, 'any', [90 0]);

AKf(12,6)
AKp(l_sh,    'm2d', 'N', 4410)
AKp(l(:,id), 'm2d', 'N', 4410, 'dr', [-30 25])
legend('N=5', 'N=20', 'N=35', 'reference', 'location', ...
    'SouthWest')
print('-dpdf', 'figure4b')
```

```
%% ---------------------------- REGULATED INVERSION
% run demo script - figure 3 is saved automatically
AKregulatedInversionDemo
```

B

The PIRATE – an anthropometric earPlug with exchangeable microphones for Individual Reliable Acquisition of Transfer functions at the Ear canal entrance

Florian Denk, Fabian Brinkmann, Alfred Stirnemann, and Birger Kollmeier (2019), *Fortschritte der Akustik –DAGA 2019*, Rostock, Germany.
(Accepted manuscript. CC-BY 4.0)

Measurements of individual Head-Related Transfer Functions and Headphone Transfer Functions require positioning a microphone at the blocked ear canal entrance. The common approach is to modify foam earplugs or silicone domes that are easily inserted into different ears to accommodate a microphone[1]. However, the soft material results in a poorly defined fit in the individual ear, as well as a poor stability and repeatability between insertions. The best option would be individually made earplugs[2], which are, however, expensive and tedious to make.

We present the open design of the PIRATE, an anthropometric ear**P**lug for **I**ndividual **R**eliable **A**cquisition of **T**ransfer functions at the **E**ar canal entrance. Its outer shape is available in 5 sizes and provides a deep, tight and reproducible fit in virtually all human ears. It was designed based on the statistical analysis of several hundred ear canal scans, as first presented in[3]. The design includes a recess to accommodate a MEMS microphone. Thus, the same microphone can be conveniently used in different earplugs without losing accuracy, and the microphone can be removed for calibration. The PIRATE or previous versions of it have been utilized in several studies with more than 200 subjects, including publicly available datasets[4,5]. We believe that it would be helpful also for other researchers, therefore the 3D models are made available to the public under a Creative Commons license at `https://doi.org/10.5281/zenodo.2574395`. From these, the earplugs can be 3D printed in sili-

[1] D. Hammershøi and H. Møller (**2005**). "Binaural technique – basic methods for recording, synthesis, and reproduction" in *Communication Acoustics*, edited by J. Blauert.

[2] V. R. Algazi, et al. (**2001**b). "The CIPIC HRTF database" in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*.

[3] A. Lindau and F. Brinkmann (**2012**). "Perceptual evaluation of headphone compensation in binaural synthesis based on non-individual recordings" *J. Audio Eng. Soc.*

[4] A. Lindau and F. Brinkmann (**2012**). "Perceptual evaluation of headphone compensation in binaural synthesis based on non-individual recordings" *J. Audio Eng. Soc.*

[5] F. Denk, et al. (**2018**). "Adapting hearing devices to the individual ear acoustics: Database and target response correction functions for various device styles" *Trends in Hearing*.

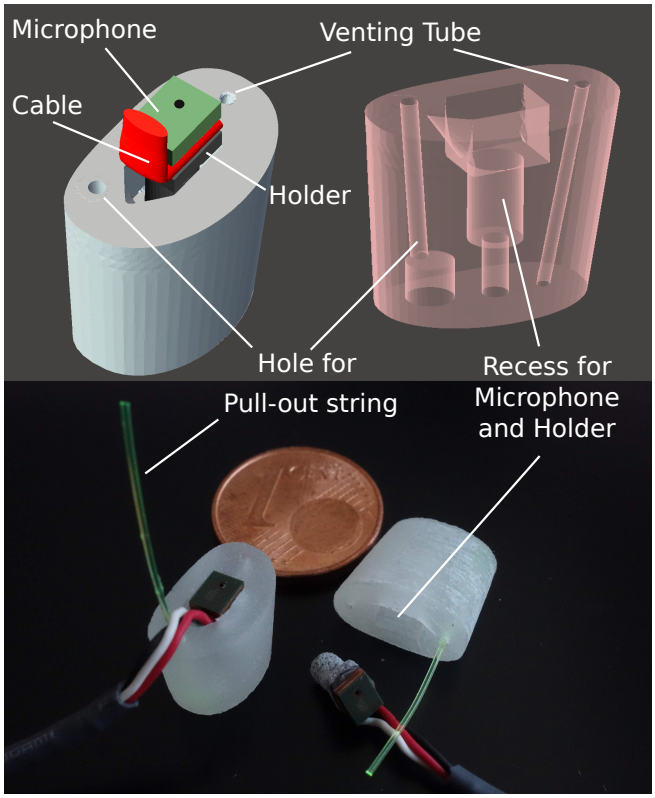


Figure B.1: CAD model (top) and photograph (bottom) of the assembled PIRATE. See text for further details. All images show right-ear, medium-sized earplugs.

cone with only minor manual working steps necessary.

We here describe the design of the PIRATE and show the achievable reproducibility of measurements in an individual human ear. Also, this document includes the technical documentation necessary to make the PIRATE from the published 3D models.

B.1 Earplug design

A CAD model of the PIRATE as well as an image of the assembled version is shown in Figure B.1. The outer form of the earplug is described in detail in the following section. The complete device consists out of the main earplug, a MEMS microphone, and a holder for the microphone (visible at the upper left). The main earplug is made out of 3D printed silicone with Shore 65 hardness, the holder was 3D printed in a standard PA plastic.

In the main earplug, several holes and recesses are included, best seen at top right of Figure B.1. The largest hole (in the middle) is a matched recess for the microphone attached to a holder. The holder is a rod (Ø = 2.4 mm) with a plate on top that matches the size of the utilized microphone (2.65 x 3.5 mm). We here utilized a MEMS microphone with a top port location, specifically a Knowles SPH1642-HT5H. Several other microphones are available with the same size and port location. MEMS were chosen over electret condenser microphones due to the superior ratio of SNR to size, better temperature stability and smaller variation between devices[6]. Also, they are a factor of about 20 cheaper than electret condenser microphones with a

[6] J. Lewis and B. Moss (**2013**). "MEMS microphones, the future of hearing aids" *Analog Dialogue*.

comparable size. See the Technical Documentation for more details on the microphone.

The recess is designed such that the microphone port is flush with the outer surface of the earplug when inserted. Also, the microphone sits very tight. A small tube connects the bottom of the recess to the inner surface of the earplug (see Figure B.1, top right). This allows to push the microphone out of the earplug without any stress on the wiring. At the bottom side of the outer face of the recess, space for the cables is provided. The design makes it very convenient to use the same microphones with different earplugs, and to take out the microphone for calibration purposes.

Also, a small venting tube (Ø = 0.7 mm) located above the microphone is included to avoid an over-pressure in the ear canal after insertion of the earplug. Due to the tight sealing, this issue had turned out to be a problem with previous versions. The diameter was chosen as small as possible and does not influence the ear acoustics.

Furthermore, the earplug features a tube where a string for pulling the earplug out of the ear can be mounted. Attempts to have such a string 3D printed with the earplug were not successful due to poor stability within acceptable diameters. Instead, a nylon string can easily be glued into the provided tube. At the inner end, the tube is broadened such that the inner end of the string can be knotted. Detailed instructions on this are given in the Technical Documentation.

B.2 Anthropometric Shape

To provide a good fit across a large variety of ear canals, the earplugs were designed in five different sizes (XS, S, M, L, XL) and in a shape inspired by the anthropometry of the human ear canal. As seen in Figure B.1, the basic form consists of two ellipsoid-like faces that form the the top and bottom of a conical body. The major and minor diameter of the ellipsoid-like faces were designed based on the major and minor diameter of the ear canal entrance at the position of the first bend, and inside the ear canal at the position of the second bend (cf. Figure B.2). For the size M earplug, the mean values extracted from 999 laser scans of human ear canals were used. The remaining sizes were designed under the assumption of normally distributed measures by calculating selected percentile values based on the mean and corresponding standard deviation (cf. Table B.2). The exact form and relative position of the faces was designed after an inspection of individual ear canal impressions. The depth of earplugs was set to 9.2 mm in all cases, because the initially favored distance between the first and second bend was too short to provide a stable microphone mount and fit of the earplug in the ear canal. This, however, was considered a negligible drawback as the outer part of the ear canal is somewhat elastic due to a layer of cartilage.

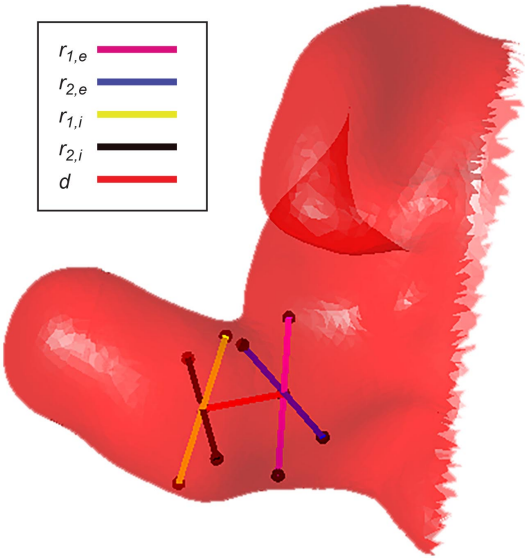


Figure B.2: Example of an ear canal scan and extracted measures. The major and minor radii at the ear canal entrance at the position of the first bend are given by $r_{1,e}$ and $r_{2,e}$; radii inside the ear canal at the position of the second bend are given by $r_{1,i}$ and $r_{2,i}$; the distance between the first and second bend is given by d.

$\mathbf{P_i}$ size	$\mathbf{P_{15}}$ XS	$\mathbf{P_{25}}$ S	$\mathbf{P_{50}}$ M	$\mathbf{P_{75}}$ L	$\mathbf{P_{85}}$ XL	σ
$r_{1,e}$	12.61	13.28	14.50	15.74	16.39	1.82
$r_{2,e}$	6.30	6.82	7.75	8.69	9.19	1.39
$r_{1,i}$	9.99	10.63	11.79	12.98	13.60	1.74
$r_{2,i}$	5.83	6.35	7.29	8.24	8.75	1.40
d	3.20	3.72	4.65	5.59	6.10	1.39

Table B.1: Selected percentile values $\mathbf{P_i}$ of ear canal measures in mm (calculated under the assumption of a normal distribution with the mean $\mathbf{P_{50}}$ and the standard deviation σ).

B.3 Reproducibility Measurements

To assess the reinsertion accuracy in measurements with the PIRATE, we conducted repeated measurements of the diffuse-field response at the blocked ear canal entrance of an individual subject. To this end, the subject was sitting in an anechoic chamber with a multi-channel 3D loudspeaker setup. Uncorrelated white noise was played from 47 uniformly distributed loudspeakers at the same time, thus generating an approximated diffuse field[7]. The PIRATE was re-inserted into the subject's ear five times, and each time 10 s of the diffuse noise recorded. The paradigm was chosen over direction-resolved measurements to rule out head movements as a source of variation[8]. The fit of the earplug in this subject's ear is shown in Figure B.3. We want to note that this measurement was attempted only once, and the subject was by no means selected, e.g., for their ear.

Figure B.4 shows the diffuse-field response at the ear canal entrance of the subject, with each line showing one repetition. The reproducibility of the measurement is excellent. Noticeable differences of a few dB only occur between 4 kHz and 10 kHz. On a closer inspection, it becomes apparent that only one measurement differs from the other four. Apart from that, no considerable differences are observed, which holds even for the high frequency range above 10 kHz. The results also demonstrate that the microphone blocks the ear canal very well and sits very deep. That is, the resonance of the cavum conchae is captured with a peak amplitude of 16 dB, which

[7] F. Denk, et al. (**2018**). "Adapting hearing devices to the individual ear acoustics: Database and target response correction functions for various device styles" *Trends in Hearing*.

[8] F. Denk, et al. (**2017**). "Controlling the head position during individual HRTF measurements and its effect on accuracy" in *Fortschritte der Akustik – DAGA 2017*.

Figure B.3: Photograph of the PIRATE inserted into a subject's ear.

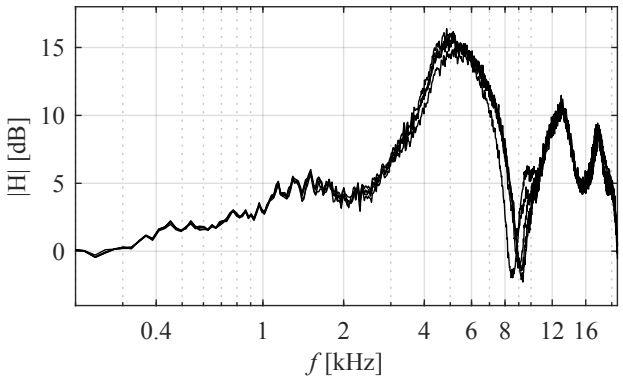


Figure B.4: Diffuse-field response of an individual subject's ear; each line represents a measurement with the earplug newly inserted.

verifies that the cavum conchae is not obstructed at all[9]. The results agree with previous measurements where a plastic ear was used to avoid variance due to positional changes of the subject[10].

[9] F. Denk, et al. (**2018**). "Adapting hearing devices to the individual ear acoustics: Database and target response correction functions for various device styles" *Trends in Hearing*.

[10] A. Lindau and F. Brinkmann (**2012**). "Perceptual evaluation of headphone compensation in binaural synthesis based on non-individual recordings" *J. Audio Eng. Soc.*

B.4 Availability

The 3D models are available at `https://doi.org/10.5281/zenodo.2574395` under the Creative Commons 4.0 CC-BY-SA license. They can be freely used, modified and redistributed, provided that the original source is attributed and modifications are published under a similar license. The models can be directly 3D printed and only some manual steps are required before using the earplugs, which are documented below.

B.5 Technical Documentation

3D Printing

Both the earplugs and the holder are designed for 3D printing. By the time this documentation was written, several 3D print shops offered the suitable technology at comparable prizes (roughly: earplug 15 €, holder: 3 €). The key aspect is 3D printing of flexible silicone with a suitable hardness. While the optimum would be at around Shore 50-55, we only found possibilities to print silicone (specifically,

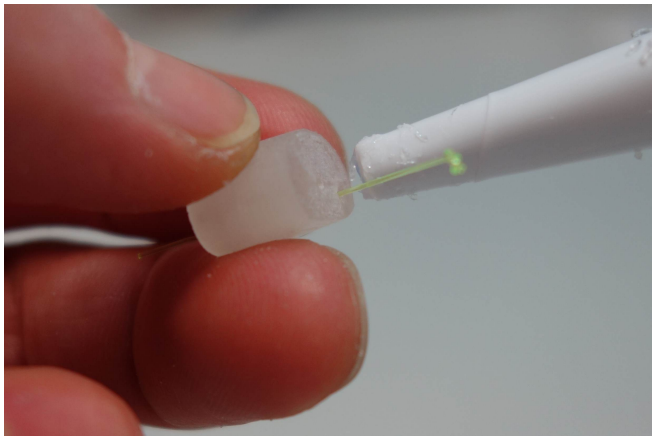


Figure B.5: Image of the pull-out string (yellow) being glued into the earplug.

silicone G1H) with a hardness of either Shore 35 oder Shore 65. The hardness should not be below Shore 50, or the earplug will be too soft to provide the defined, reproducible fit that is intended by design. We made good experiences with the Shore 65 silicone. For the holder, any rigid plastic material can be used. However, the resolution should be at least 200 μm.

Preparing the Earplug

Three manual steps are required before the earplug can be used.
First, clean the earplug from residual dust-like material that may stick to the surface after 3D printing.
Second, rework the venting tube and the hole for the pull-out string. Especially the venting tube may be non-permeable after 3D printing, although it should be well-defined. Stick something with suitable diameter ($<$ 1 mm) through it to fix this. Any tiny drill or a stiff piece of wire with a flat end does the job, needles are not recommended.
Third, insert the pull-out string for safely removing the earplug from the subjects' ears. We recommend a mono-filament fishing line with a diameter of 0.4 mm or more. Cut a piece of about 10 cm and stick it through the appropriate hole (the one that ends in a larger cavity at the inner end, see also Figure B.1). Then, make a knot into the inner end, and make sure this knot fits into the cavity. Push back the knot from the cavity by the length of the earplug, apply glue into the knot cavity (as shown in Figure B.5) and pull the knot into its cavity. The glue should stay flexible after it dried – we recommend silicone sealant. Finally, cut the pull-out string to length. We recommend to shorten it to about 1 cm and use tweezers to remove the earplug.

Microphone Assembly

MEMS microphones are usually only available without cables attached. The wires should be soldered as flat as possible on their pads, and guided away at one of the short sides (see Figure B.6). Since the port location is asymmetric, make sure to do this consistently in all your microphones. In the models, 0.9 mm are accounted for the wiring and glue layer between the holder and the micro-

phone. If you need more or less space, you are advised to modify the 3D model of the holder instead of grinding away material or attaching extra glue.

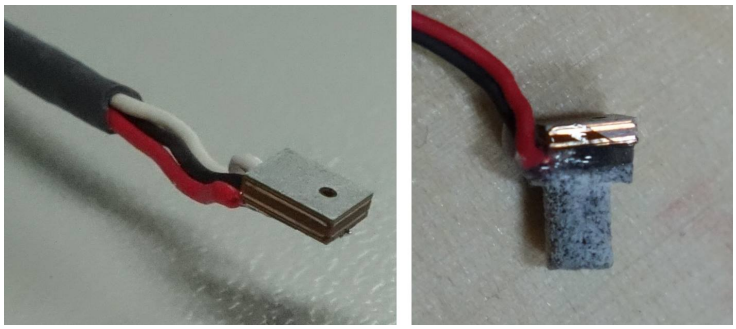


Figure B.6: Microphone, with wires attached (left) and glued onto the holder (right).

To glue the microphone to the holder, we recommend fast-drying epoxy raisin. Make sure that the microphone is level with the holder. Avoid accumulations of glue at the sides of the microphone, or it may not fit into the recess. Also, guide the cables towards the top surface of the microphone using some glue. A good example how it should look like at the end is shown at the right side of Figure B.6.

Microphone and Power Supply

The microphone we utilized (Knowles SPH1642-HT5H) provides an SNR of 65 dB, has a sensitivity of -38 dB V/Pa and cost less than 1.5 € by the time this documentation was written. It requires a power supply between 1.5 and 3.5 V, and we highly recommend to utilize an amplifier between the microphone and the AD converter. Generally, the necessary circuitry is identical to electret condenser microphones.

Practical Tips

Finally, here are some practical tips for using the PIRATE based on our experience:

- Make sure to insert the correct earplug side with the right orientation. The cable guide of the recess always points downwards. For distinguishing left and right earplugs: The more rounded side of the earplug goes to the front, the straight one to the back. The pull-out string is located more towards the concha, not to the front (see also Figure B.3).
- Check the ear for large accumulations of cerumen before inserting the earplugs. Cerumen will be pushed deep into the ear canal, which might occasionally clog the ear.
- Choose the earplug size before you put in the microphones by test-inserting them into the ear. If an earplug sits loosely, try a larger size.
- The earplug should sit tight and firmly, with its outer surface flush with the ear canal entrance. This position is the easiest to reproduce, and the reference position for most measurements.

- The inner end usually sits at the second bend of the ear canal, therefore be gentle when pushing it in or you might hurt your subjects. In some ear canals, this will result in the earplug being tilted against the ear canal entrance plane. In this case, try to make the rear edge of the earplug (sticks out further) flush with the ear canal entrance.
- Do not pull at the microphone cable, use the pull-out string. To relief stress on the cable, it is a good idea to tape the cable to the cheek or neck of your subject using medical tape.

Last but not least: we are happy about feedback on any experiences with the earplugs, positive or negative!