TECHNISCHE UNIVERSITÄT MÜNCHEN
Lehrstuhl für Mensch-Maschine-Kommunikation
Arbeitsgruppe Technische Akustik

Interrelations of Virtual Acoustics and Hearing Research by the Example of Binaural Synthesis

Florian Völk

Vollständiger Abdruck der von der Fakultät für Elektrotechnik und Informationstechnik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktor-Ingenieurs (Dr.-Ing.)

genehmigten Dissertation.

Vorsitzender:	Univ.-Prof. Dr.-Ing. habil. Dr. h.c. A. W. Koch
Prüfer der Dissertation:	1. apl. Prof. Dr.-Ing., Dr.-Ing. habil. H. Fastl
	2. Univ.-Prof. Dr.-Ing. W. Hemmert

Die Dissertation wurde am 15. November 2012 bei der Technischen Universität München eingereicht und durch die Fakultät für Elektrotechnik und Informationstechnik am 06. Mai 2013 angenommen.

Meinen Eltern.

Acknowledgments

At this point, I wish to express my gratitude to all the people who supported me up until now, especially to those who contributed to this thesis.

My first thanks are due to my teacher and mentor Prof. Dr.-Ing. Hugo Fastl, who gave me the chance to work on this thesis and various other projects as well as to teach as a member of his group *AG Technische Akustik* at the *Institute for Human-Machine Communication* of *Technische Universität München.* Without his support, especially in raising funds, this work would have never been started or finished, and I might not have found this exciting profession, allowing me to combine personal interests and daily work.

Additional thanks go to the second examiner Prof. Dr.-Ing. Werner Hemmert and to Prof. Dr.-Ing. Alexander W. Koch, chairman of the committee, for taking over these responsibilities. I am thankful to the head of the *Institute for Human-Machine Communication*, Prof. Dr.-Ing. Gerhard Rigoll, for his support through the course of this thesis and to *Deutsche Forschungsgemeinschaft (DFG)*, who supported major parts of this work by grant FA 140/4.

My colleagues of the *AG Technische Akustik*, Dr.-Ing. Markus Fruhmann and Dr.-Ing. Stefan Kerber, complemented by Dipl.-Ing. Jakob Putner, introduced me to every facet of the *Munich School of Psychoacoustics.* Without their guidance, I might never have discovered interest in the field. My peer Dr.-Ing. Daniel Menzel, whom I shared an office with for the majority of my time at the institute, contributed substantially to this thesis by providing a comfortable working atmosphere in the office and at various conferences, as a sedulous discussion partner, and especially as patient expert listener, participating in countless psychoacoustic experiments.

This work was substantially supported by my students, especially in the context of thirteen Masters and fourteen Bachelors theses as well as three scientific assistant jobs. It is my pleasure to thank Dipl.-Ing. Fabian Heinemann, Dipl.-Ing. Josef Konradl, Dipl.-Ing. Elias Faccinelli, Dipl.-Ing. Thomas Musialik, M. Sc., Dipl.-Ing. Daniela Petri, Dipl.-Ing. Andreas Dunstmair, Dipl.-Ing. Tobias Riesenweber, Dipl.-Ing. Christian Landsiedel, Dipl.-Ing. Maximilian Schmidhuber, Dipl.-Ing. Armas Gottschalk, Dipl.-Ing. Timon Seidl, Dipl.-Ing. Ulrich Mühlbauer, Dipl.-Ing. Martin Straubinger, Dipl.-Ing. Luis Roalter, Dipl.-Ing. Florian Lindner, Dipl.-Ing. Tanja Schulze, and Dipl.-Ing. Thomas Brandl, as well as Ralph Truckenbrodt, B. Sc., Andreea Fintoc, B. Sc., cand.-Ing. Edoardo Hager, and cand.-Ing. Yuliya Fedchenko. Every single one of them contributed in their own way, and each contribution was special to me.

Thanks also to the colleagues of other groups at the *Institute for Human-Machine Communication* for the comfortable atmosphere. Special thanks are due to Dr.-Ing. Dejan Arsić, Heiner Hundhammer, Ernst Ertl, Gertrud Günther, and Martina Römpp for their recurring support in so many different ways. I enjoyed the technically high valued and morally vital encouragement of the “Munich Acoustic Communication Circle”, namely

Dr.-Ing. Markus Mummert, Dipl.-Ing. Tilman Horn, and Dr.-Ing. Claus von Rücker. Many ideas discussed in this thesis were born or raised at our gatherings.

Without the highly appreciated support of Prof. Dr.-Ing. Manfred Zollner (University of Applied Sciences Regensburg), Prof. Dr.-Ing. Helmut Fleischer (Universität der Bundeswehr München), Prof. Dr.-Ing. Jürgen Peissig (Sennheiser electronic GmbH & Co. KG), Dipl.-Ing. Sebastian Goossens (Institut für Rundfunktechnik), and Dipl.-Ing. Thomas Steinbrecher (Bose GmbH), this work would not have been possible in its present form.

Finally, special thanks for their time and efforts are due to Dr.-Ing. Daniel Menzel, Dipl.-Ing. Jakob Putner, Dr.-Ing. Dejan Arsić, Dr.-Ing. Stefan Kerber, StR Carolin Johanna Völk, and Charlene Renée Nance-Mannon, M. L. A., M. Ed., who contributed significantly by valuable comments and fruitful discussions on earlier versions of the manuscript.

My deep-rooted gratefulness for all the encouragement not only while I have been working on this thesis but also during the remainder of the past 32 years belongs to my parents Mathilde and Georg and my sister Carolin. Without their patience and unselfish support, it would have been impossible for me to accomplish what I have reached so far, in my personal as well as my professional life.

Thank you so much!

Munich,
September 2012

Florian Völk

Contents

1 Introduction

The term *virtual acoustics* as referred to within this thesis summarizes the group of sound reproduction and synthesis approaches attempting to elicit the hearing sensations occurring in the real or hypothetical scenario to be simulated. As such a superordinate concept, virtual acoustics can be and has already proven useful for high-quality audio reproduction purposes in a variety of applications, covering different domains of everyday life, commercial products, and scientific work. Typical examples include teleconferencing (Begault 1999, de Bruijn et al. 2000, Boone and de Bruijn 2003), navigating and alerting for the blind (Loomis et al. 1998), as well as warning and guidance support in traffic situations, for example in vehicles (Cohen et al. 2006, Poitschke et al. 2009) or aviation (Begault 1998, Begault et al. 2006).

Virtual acoustics by definition exactly fits the requirements of playback methods for various multimedia applications. Among them are audio systems for car entertainment (Farina and Ugolotti 1999, Jones et al. 2007, Bai and Lee 2010), home cinema (Theile et al. 2002, Klehs and Sporer 2003), and venues like movie theaters (Sporer and Klehs 2004), concert halls (de Vries 1996, Pellegrini and Kuhn 2005), and auditoria (Moldrzyk et al. 2007). The increasing popularity of mobile communication and audio playback devices triggers the development of compatible virtual acoustics systems (Pörschmann 2002, Huopaniemi 2007). Possibly the most obvious application scenario of virtual acoustics is the employment in virtual and augmented reality setups like Cave Automatic Virtual Environments (Hammershøi and Sandvad 1997, Melchior et al. 2006) or driving simulators (Völk et al. 2007, Menzel et al. 2011b). Scientific fields of application cover all kinds of *hearing research* such as Psychoacoustics (Blauert et al. 2000, Zahorik 2002, Sivonen 2007, Völk 2009), Audiology and medical research (Kayser et al. 2009, Völk and Fastl 2010, Völk et al. 2012a), as well as sound quality assessment (Völk 2012b). The latter includes applications in room acoustics (Lokki 2005), vehicle sound design (Otto 1997, Farina and Ugolotti 1998, Gallo et al. 2010), and the development of audio reproduction systems like loudspeaker boxes (Hiekkanen et al. 2009) or car entertainment systems (Granier et al. 1996, Farina and Ugolotti 1997).

The goal of triggering hearing sensations in a perfectly controlled way is rather ambitious especially since a hearing sensation typically depends on different physical parameters (Zwicker and Feldtkeller 1967, p. 43). Therefore, it is usually impossible to synthetically generate all relevant physical cues as they were in the target situation, particularly since the relevance of different physical cues resembles an open question unlikely to be solvable in general (Macpherson and Middlebrooks 2002, Seeber 2007, 2010). Thus, it is inevitable for the majority of applications to reduce a virtual acoustics system's performance requirements to a feasible extent (Wightman et al. 1992). Similar to the dichotomous movement of audio coding (Fastl 2010) towards low data rates on the one hand (Herre et al. 1994, Herre 2004) and extraordinary high data rates on the other

(Funasaka and Suzuki 1997, Suzuki 2004), a tendency is visible for virtual acoustics to be implemented either with strictly limited performance requirements (Davis and Fellers 1997, Pellegrini and Horbach 2002, Faller 2004) or to reach the highest achievable quality (Berkhout et al. 1993, Hawksford 1997b, Klehs and Sporer 2003). However, the ideal approach would reduce the performance requirements in such a way that the resulting transmission errors are not or at least unlikely detectable by human listeners. While this aim is obvious, purposeful listening experiment procedures and results applicable to derive meaningful performance requirements for virtual acoustics systems are largely missing, as stated in the conclusions of Spors (2005).

1.1 Motivation and Objective

In 1967, Eberhard Zwicker and Richard Feldtkeller, two of the German prime fathers of Psychoacoustics, recommended in the introduction to their book *Das Ohr als Nachrichtenempfänger* (*The ear as a communication receiver*, Zwicker and Feldtkeller 1967, pp. vi–vii; English edition 1999, p. xvii) that

> *"in the design of reproduction systems for speech and music* [. . .] *it is* [. . .] *important to learn what degradations of the reproduced signal go unnoticed by the ear and to use this knowledge in the specification of the system's performance requirements."*

This recommendation inherently assumes the *motivation* that a meaningful audio reproduction system must fulfill requirements specified by the characteristics of the final receiver, the human hearing system. The reproduction system design process according to Zwicker and Feldtkeller consequently starts not by figuring out what can be reached technically with the least effort (without knowing or being aware of the perceptual consequences), but with the identification and definition of the system specifications that are optimally matched to the final receiver, the human hearing system. This procedure seems worthwhile from an engineering point of view, especially considering that the specific receiver discussed here cannot be modified by the system designer (cf. also Zwicker and Zwicker 1991). Based on the optimal specification, the performance of the system can be degraded if necessary or desired, but in a controlled way; controlled insofar as the perceptual consequences of the degradation are known and accepted deliberately.

The typical design procedure of audio reproduction and signal processing systems based on Zwicker and Feldtkeller's method is accordingly divided in two steps:

1. *Determination of the just noticeable degradations* of the reproduced signals by perceptual comparison of degraded versions to the originals.

2. *Specification of system performance requirements* based on the just noticeable degradations and a cost-benefit analysis of the attempted application.

Presumably the most prominent technology in the field of audio signal processing developed based on Zwicker and Feldtkeller's method is the perceptual audio coding, colloquially referred to as MP3-coding (Blauert and Tritthart 1975, Theile et al. 1987, Brandenburg

1987), encountered in various situations of current daily life. The development of perceptual audio coding was possible only based on extensive studies of the signal degradations just noticeable by the human hearing system (Zwicker 2000), resulting especially in the concepts of the critical bands (Fletcher and Munson 1937, Zwicker et al. 1957, Zwicker and Fastl 1972), masking patterns (Fletcher and Munson 1937, Zwicker 1965, 1975, Fastl 1975), and loudness (Fletcher and Steinberg 1924, Zwicker and Feldtkeller 1955, Zwicker 1958). Early motivations and procedures to adapt audio signal processing to the human hearing system are reported by Stevens and Davis (1938) or Feldtkeller and Zwicker (1956). Applied to virtual acoustics, a field of technology far from realization in its present incarnations back in Zwicker and Feldtkeller's days, their method still represents the preferable approach to system design (e. g. Seeber 2002b, Seeber and Fastl 2003).

It is the *objective* of this thesis to deduce a unified framework permitting the application of Zwicker and Feldtkeller's method for the design of audio reproduction and signal processing systems to virtual acoustics. This includes a system-theoretically thorough baseline system, serving not only as the reference situation for determining just noticeable degradations using psychoacoustic methods, but also providing a high quality stimulus presentation system for hearing research in general. By reducing the idealized system-theoretically optimal procedure taking into account just noticeable degradations, a perceptually motivated engineering-approach to virtual acoustics is derived. The resulting framework allows for the development of the technically least costly implementations while keeping degradations unrecognizable or at least at a controlled amount. Regarding the application of virtual acoustics to the audio playback in hearing research, the stimulus definition and associated requirements are revised. Furthermore, auditory-adapted audio signal processing and analysis methods as well as perceptual system analysis procedures are proposed, providing the methodical basis for the discussion. Motivated by a theoretical revision and classification of existing approaches to virtual acoustics, binaural synthesis is selected as the exemplary procedure to be discussed in detail, regarding its system-theoretic basis and specific implementation factors, using the proposed framework. As a result, the just noticeable degradations are defined as generally applicable as possible, and the suitability of binaural synthesis for the audio playback in hearing research is addressed. The validity of the introduced engineering approach, its results, and the applicability in hearing research is verified by exemplary binaural synthesis implementations, which are in turn evaluated using psychoacoustic methods.

This work is intended to provide an integrated and comprehensive theoretical framework for the auditory-adapted generation and for the instrumental and perceptual evaluation of virtual acoustics by the example of binaural synthesis. The framework is developed as a common and well-defined basis for the application of virtual acoustics in research and development as well as for commercial projects. Setting up a virtual acoustics system accordingly avoids on the one hand common implementation errors while shortening the setup time, and helps on the other hand making the psychoacoustic data acquired with virtual acoustics as the playback method more reliable, more reproducible, and more generally applicable. Further, the framework allows for a global and structured view on the results of the variety of studies addressing perceptual properties of virtual acoustics systems by providing a common and thorough theoretical basis.

1.2 Structure and Overview

The body of this thesis is organized in five chapters, complemented by the introduction, a concluding summary, and a mathematical and system-theoretic appendix. The chapters are intended to cover the respective topic and may therefore be studied as single units. However, the presentation order is selected so that later chapters refer back to definitions, methods, and results discussed earlier, pointing out relations between the chapters, especially with regard to the overall statement of this thesis. At those points, it may be helpful to go back and look up the respective data using the given cross-references. Where necessary, clearly marked definitions of terms not treated consistently or not discussed in the established literature state explicitly their application within this work.

In the remainder of this section, the five main chapters are summarized regarding general scope and major results, clarifying the structure of the presentation, and giving readers with limited time a condensed overview of the thesis. Those with a deeper interest are referred to the concluding summaries and the detailed discussions within the chapters.

Chapter 2: Basic Considerations, Methods, and Terminology Initially, basic methods and terms concerning the system-theoretic and psychoacoustic procedures applied in this work are discussed and defined, especially resulting in the following achievements:

- *Refined formulae for the frequency dependence of critical bandwidth and critical-band rate*, implementing the critical-band concept, a model of the auditory frequency resolution: The proposed formulae fit the respective listening experiment results more accurately than current formulae and are better suited for the direct application in digital signal processing by implementing a bandwidth converging to zero at $f = 0\,\mathrm{Hz}$ and an invertible critical-band rate function.
- Psychoacoustically motivated tools for the instrumental and perceptual assessment of audio signals and transmission systems:
  - *Auditory-adapted analysis* (AAA) including an auditory-adapted spectrogram representation combining relevant magnitude and phase information.
  - *Loudness transfer functions* (LTFs).
  - *Quality assessment by just noticeable sound changes.*

In combination, the tools and definitions introduced in chapter 2 provide the methodical basis for a psychoacoustically motivated and instrumentally supported discussion of virtual acoustics and audio signal processing in general in the following chapters.

Chapter 3: Conjunctions of Virtual Acoustics and Hearing Research In this chapter, a *categorization of approaches* for the generation of virtual acoustical environments regarding the respective functional principle is proposed. Approaches to virtual acoustics are considered either physically or psychoacoustically motivated.

On that basis, a global overview of approaches to virtual acoustics covering established systems and the current state of the art is given, resulting in the *selection of the approach* *to be discussed* primarily within this thesis: binaural synthesis. This physically motivated approach is considered most promising and best suited for the audio playback in hearing research, in contrast to psychoacoustically motivated approaches.

Regarding the application of *virtual acoustics as the playback method in hearing research*, psychoacoustic methods and the stimulus definition are reviewed with a focus on the playback procedure and on aspects of the statistical analysis applied here. A discussion of *issues in conventional headphone reproduction and related equalization procedures* provides a link between traditional playback methods and the framework introduced in this work. Thereby, erroneous assumptions frequently associated with free-field equalized headphone playback in hearing research are identified and clarified, resulting in a *revision of the application range for the free-field equalization.* In the course of this discussion, a common lack of a nonindividual stimulus definition in conventional headphone reproduction regardless of the equalization procedure is identified and traced back to the psychoacoustically motivated fundamental concept.

A reflection on the application of *psychoacoustic methods for the quality evaluation of virtual acoustics systems* concludes this chapter, motivating and introducing a modified stimulus definition for this specific application of psychoacoustic methods.

Chapter 4: Theoretical Aspects of Idealized Binaural Synthesis Using the methods and tools introduced in the previous chapters, the *system-theoretic background* of static binaural synthesis with

- probe microphone recording, positioning the probe tube tips close to the eardrums,
- blocked auditory canal entrance miniature microphone recording, and
- artificial head recording

is revised in chapter 4 for human and artificial head playback. Particularly, the *identification of theoretical shortcomings* such as assumptions and approximations applied for the derivation of static binaural synthesis is addressed, providing the basis for a detailed discussion of implementation factors, dynamic binaural synthesis, and results in chapter 5. The major outcome of chapter 4 is a theoretic framework for binaural synthesis, allowing for the well-founded selection and exact specification of the implementation employed in an application scenario, which is crucial to allow for the meaningful discussion of listening experiment results acquired using binaural synthesis.

Further, a *binaural synthesis quality criterion* (BSQC) is introduced based on the artificial head validation of binaural synthesis transfer functions, providing an authenticity measure for the ear signals generated by binaural synthesis with artificial head playback. Combined with the introduced system-theoretic framework, the BSQC allows for identifying and tracking down binaural synthesis errors to the causative partial system.

In summary, the *possibilities and limitations of the different recording methods* (probe microphone, blocked auditory canal, and artificial head recording) are addressed in this chapter on a system-theoretic basis for binaural synthesis with human head and artificial head playback. In the final section, the results are discussed with regard to synthesizing authentic ear signals.

Chapter 5: Practical Aspects of Applied Binaural Synthesis This chapter primarily contains a discussion of the perceptually motivated implementation and parameterization of *dynamic binaural synthesis*. Special attention is paid to the perceptual consequences of deviations between feasible systems and the requirements and assumptions derived and identified by the system-theoretic discussion in chapter 4. Thereby, the implementation of dynamic binaural synthesis using the methods of the linear dynamic system theory is discussed, especially including aspects of the signal processing and the discrete measurement grid. For determining the *grid resolution* necessary for transparent binaural synthesis, a listening experiment based method is proposed and evaluated. Further, interpolation procedures to increase the grid resolution are reviewed.

The *inter- and intra-individual variability* of the headphone and recording situation transfer functions are addressed in detail, for human and artificial head recording, including aspects of the headphone production spread. Furthermore, possibilities and limitations of probe microphone ear signal measurement are discussed along with the resulting variability.

As a tool supporting the implementation of binaural synthesis especially with regard to inversion problems, *auditory-adapted exponential transfer function smoothing* (AAS) is proposed. The procedure is derived system theoretically as well as parameterized and evaluated by listening experiments.

Combining the theoretical foundation given in chapter 4 with the headphone transfer function properties discussed in this chapter, the necessity for selecting appropriate headphones when implementing binaural synthesis based on blocked auditory canal recording is shown. Addressing this issue, a *blocked auditory canal headphone selection criterion* (HPSC) is proposed, evaluated, and related to the binaural synthesis quality criterion introduced in chapter 4.

The chapter is concluded by a series of *loudness comparison experiments* between a real loudspeaker and its binaurally synthesized counterpart, representing the overall evaluation of the framework and of the binaural synthesis procedures introduced in this thesis. The results show the applicability of the framework and indicate that the loudness can be reproduced correctly by static and dynamic binaural synthesis with individual recording and individual magnitude and phase equalization only if headphones are used which fulfill the blocked auditory canal headphone selection criterion formulated here. Nonindividual measurement, nonindividual equalization, and inappropriate headphones are most likely to result in loudness deviations.

In addition, an *explanation of the case of the missing 6 dB* is derived from the experimental results: Identical sound pressure time functions in the auditory canals ensure the same loudness in loudspeaker and headphone reproduction if the ambient conditions are kept constant. In contrast, equal loudness is not necessarily ensured by equal root-mean-square sound pressure levels in the canals.

Finally, a *schematic working model of auditory localization, loudness, and sound color perception* is derived to summarize the results of the loudness adjustment experiments. The working model represents an extension of Theile's association principle and is shown to account for listening experiment results regarding loudness and localization.

2 Basic Considerations, Methods, and Terminology

This fundamental chapter includes a clarification of terms, the definition of variables, and the discussion of methods and procedures employed. Especially, *refined formulae for the dependence of critical bandwidth and critical-band rate on frequency*, *auditory-adapted analysis*, and *loudness transfer functions* are introduced. Supplemented by *just noticeable sound changes regarding auditory localization*, these psychoacoustically motivated tools allow for the instrumental and perceptual assessment of audio signals and systems.

2.1 Fundamental Variables and Conventions

In the context of this thesis, lower case variables denote time dependent signals or impulse responses (IRs). Being t the independent variable representing time, the analog voltage variable $u\,(t)$ is abbreviated for example as u without explicit notation of the time dependence. Upper case letters represent complex Fourier spectra or transfer functions (TFs), for example $U\,(f) = \mathcal{F}\{u\,(t)\}$ with the temporal frequency f, shortened to U. It is necessary at some points to denote whether a signal is represented digital or analog. While analog signals are signified with letters indicating the specific signal's physical nature, digital time domain sequences are, with the sample index n, always denoted by $s\,[n]$ and abbreviated as s. Different sequences are distinguished by additional indices. With the frequency bin index k, the discrete Fourier transform of the sequence s is given by $S\,[k] = \mathcal{F}_{\mathrm{D}}\{s\,[n]\}$, shortened to S. Linear time-invariant (LTI) systems are represented in the time domain by their IR h and in the frequency domain by the corresponding TF H. Different systems are indicated by additional indices. The variables frequently used in this work are summarized in appendix B.

While lower case subscripts differentiate signals and systems, additional upper case indices denote the left (L) or right (R) channel of a symmetric two channel system or a specific channel of an arbitrary multi-channel setup, if necessary. A major part of the following treatment examines a transmission system's left and right channels or the single channels of multi-channel systems with more than two channels in a similar manner. Therefore, variables representing the same components for all channels are summed up as vectors and consequently set in bold fonts, for example

$$\mathbf{U} = \begin{pmatrix} U_{\mathrm{L}} \\ U_{\mathrm{R}} \end{pmatrix}. \tag{2.1}$$

If not denoted otherwise, the division or multiplication of two vectors of the same size is used as shorthand for element-wise division or multiplication, meaning for example the element representing the left side of the first vector is divided or multiplied by the corresponding element of the second vector. A prerequisite for the division of two spectra or TFs is a dividend with non-zero magnitude, taken for granted for all calculations in this thesis including the invertibility of IRs. Possible problems concerning these issues in practical implementations have been discussed theoretically and can be reduced by adequate preprocessing (Neely and Allen 1979, Kirkeby et al. 1998, Kirkeby and Nelson 1999, Norcross et al. 2004b). A theoretically thorough, auditory-adapted, and practically proven preprocessing method referred to as auditory-adapted exponential transfer function smoothing (AAS) is introduced in section 5.1.5 (cf. also section 5.2.3).

Object positions are represented within the work on hand by the corresponding position vectors $\mathbf{x}$ in an arbitrarily chosen coordinate system. The specific object is indicated by subscripts, if necessary. A position vector is composed of x, y, and z, symbolically denoting the object position, as well as r_x, r_y, and r_z, indicating the object orientation (the rotations around the axes). Here, the loudspeaker (LS) position vector $\mathbf{x}_{\text{ls}} = (x, y, z, r_x, r_y, r_z)^T$ is given as an example. If multiple corresponding objects are to be addressed, their position vectors are combined to a matrix. The position vector $\mathbf{x}_{\text{h}} = (x, y, z, r_x, r_y, r_z, \delta)^T$ denotes head position and orientation, where the parameter δ takes into account symbolically the rotations of the listener's head and torso with respect to the other body.

For the discussion and categorization of approaches to virtual acoustics, a definition of virtual acoustics itself is required. Definition 1 shows the formulation used in this thesis.

Definition 1 (*Virtual Acoustics and the Reference Scene*)

The term virtual acoustics describes the audio reproduction or synthesis methods and procedures aiming at eliciting the hearing sensations occurring in the real or hypothetical scenario to be simulated, the reference scene.

In the context of virtual acoustical scenarios and auditory localization in general, head-related coordinates have proven helpful (Blauert 1997, figure 1.4). Here, a Cartesian coordinate system centered at the midpoint of the interaural axis connecting the upper edges of the auditory canal entrances is used (cf. figure 2.1 and definition 2).

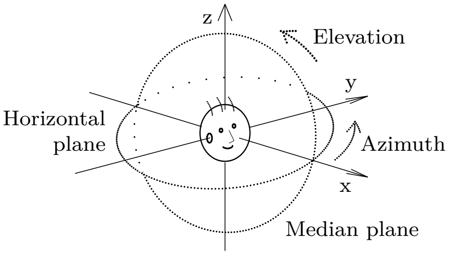

Figure 2.1: Head-related Cartesian coordinate system used in this thesis. The coordinate axes define the orthogonal horizontal, median, and frontal planes intersecting at the origin, as given by definition 2. Shown are the horizontal and median planes with the corresponding azimuth and elevation angles.

Definition 2 (*Interaural Axis, Horizontal, Frontal, and Median Plane*)

The horizontal plane is defined by the lower edges of the eye sockets and the upper edges of the auditory canal entrances. The latter points also define the interaural axis. Horizontal, frontal, and median plane are orthogonal and intersect at the interaural axis' center, while the frontal plane also includes the interaural axis.

The coordinate system shown by figure 2.1 defines three orthogonal planes (Terhardt 1998, p. 223): the horizontal plane including the interaural axis and the lower edges of the eye sockets, the frontal plane also including the interaural axis, and the median plane defined by the origin and the orthogonality requirement. A spherical head-related coordinate system with the origin at the center of the interaural axis is employed in addition, defining the azimuth angle counterclockwise in the horizontal plane with respect to the median plane and the elevation angle in the median plane with respect to the horizontal plane (cf. figure 2.1). Positive angles indicate elevations above, negative angles below the horizontal plane. The radius coordinate represents the distance to the origin.

2.2 Transfer Function Definitions

Throughout this work, homogeneous and inhomogeneous LTI systems are considered and systems are defined between different physical magnitudes (cf. Terhardt 1998, pp. 61–63). The typical homogeneous electric TF is given between two voltage spectra, while the most common homogeneous TF in acoustics relates two sound pressure spectra. An inhomogeneous TF can for example be defined between a sound pressure and a voltage spectrum, describing a linear electroacoustic transducer (Terhardt 1998, p. 63). In order to relate the spectra of physical signals to their digital representations, TFs are employed here symbolically, too. Assuming 24 Bit word-length and the sampling theorem (Nyquist 1924, Shannon 1949) being fulfilled, digital to analog (D/A) and analog to digital (A/D) conversion can be regarded linear and invertible within the audible frequency and dynamic range (Kammeyer and Kroschel 2006, p. 10). This is given for signals in the audible frequency range when using high quality audio interfaces (cf. e. g. RME – Intelligent Audio Solutions 2011). Multiple-port systems are fully described only if the potential magnitude P and the flow magnitude Q are known at each port (Terhardt 1998, pp. 75–78). Consequently, four TFs can be given for the two-port system

$$\begin{pmatrix} P_1 \\ Q_1 \end{pmatrix} = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} P_2 \\ Q_2 \end{pmatrix} \tag{2.2}$$

with the input spectra P_1 and Q_1 and the output spectra P_2 and Q_2. Assuming a typical situation with the output port loaded by the complex, frequency dependent impedance Z_L and the input port connected to a potential source with the complex frequency dependent inner impedance Z_Q and inner potential P_i, the TF

$$H_\mathrm{PP} = \frac{P_2}{P_\mathrm{i}} = \frac{Z_\mathrm{L}}{Z_\mathrm{Q}\left(A_{21}Z_\mathrm{L} + A_{22}\right) + A_{11}Z_\mathrm{L} + A_{12}} \tag{2.3}$$

relates the output potential at the two-port system to the inner potential of the driving source (Terhardt 1998, p. 78). This relation depends on the source and load impedances. H_PP represents the potential transformation of the system as long as the source and load impedances remain constant and the system is not modified, which holds true also for the TF relating Q_2 and P_i, while the TFs relating P_2 to Q_1 and Q_2 to Q_1 only depend on the load impedance Z_L. The TF between the potential magnitudes describes an LTI system as long as the system including source and load impedances is not modified. A single TF is not a valid system description if impedances or system characteristics vary.

If not denoted otherwise, pressure sensitive microphones are used here (cf. Zollner and Zwicker 1993, p. 182), for convenience referred to as microphones. Their dynamic ranges are assumed to allow for measurements without audible noise, while the distortion factors play a minor role for TF measurements if a procedure is employed ensuring nonlinear distortions not to corrupt the results (Völk et al. 2009b). The microphone TFs

$$\mathbf{H}_\mathrm{mic} = \frac{\mathbf{U}_\mathrm{mic}}{\mathbf{P}_\mathrm{mic}} \tag{2.4}$$

are defined here between the sound pressure spectra at the microphones $\mathbf{P}_\mathrm{mic}$ and the microphone output voltage spectra $\mathbf{U}_\mathrm{mic}$. If no explicit distinction is required, the subscript *mic* is used. Further, probe microphone recordings at the eardrums (subscript *pm*), miniature microphone measurements at the blocked auditory canal entrances (*m*), and recordings with artificial head microphones (*ahm*) are of interest here. The TFs

$$\mathbf{H}_\mathrm{i} = \frac{\mathbf{S}_\mathrm{mic}}{\mathbf{U}_\mathrm{mic}} = \mathbf{H}_\mathrm{am} \cdot \mathbf{H}_\mathrm{ad} = \frac{\mathbf{U}_\mathrm{ad}}{\mathbf{U}_\mathrm{mic}} \cdot \frac{\mathbf{S}_\mathrm{mic}}{\mathbf{U}_\mathrm{ad}} \tag{2.5}$$

describe the audio input systems transforming the microphone output voltages $\mathbf{u}_\mathrm{mic}$ to the digital sample sequences $\mathbf{s}_\mathrm{mic}$. The input systems are a combination of the microphone pre-amplifiers (*am*) and the A/D-converters (*ad*). With equations 2.4 and 2.5, the sound pressure spectra at the microphones are formulated by

$$\mathbf{P}_\mathrm{mic} = \frac{\mathbf{U}_\mathrm{mic}}{\mathbf{H}_\mathrm{mic}} = \frac{\mathbf{S}_\mathrm{mic}}{\mathbf{H}_\mathrm{mic} \cdot \mathbf{H}_\mathrm{i}} = \frac{\mathbf{S}_\mathrm{mic}}{\mathbf{H}_\mathrm{mic} \cdot \mathbf{H}_\mathrm{am} \cdot \mathbf{H}_\mathrm{ad}}. \tag{2.6}$$

In a typical LS playback scenario, the output voltage u_da produced by an audio interface from the sequence s_ls drives an LS amplifier (*als*) providing the voltage u_ls to the LS. This is represented in the frequency domain by the TFs

$$H_\mathrm{da} = \frac{U_\mathrm{da}}{S_\mathrm{ls}}, \quad H_\mathrm{als} = \frac{U_\mathrm{ls}}{U_\mathrm{da}}, \text{ and } \quad H_\mathrm{o} = \frac{U_\mathrm{ls}}{S_\mathrm{ls}} = H_\mathrm{da} \cdot H_\mathrm{als}. \tag{2.7}$$

If headphones (HPs) are used for audio playback, HP amplifiers (*ahp*) driven by input voltages $\mathbf{u}_\mathrm{da}$ produced by an audio interface from the sequences $\mathbf{s}_\mathrm{hp}$ provide the HP input voltages $\mathbf{u}_\mathrm{hp}$. The corresponding TFs are given by

$$\mathbf{H}_\mathrm{da} = \frac{\mathbf{U}_\mathrm{da}}{\mathbf{S}_\mathrm{hp}}, \quad \mathbf{H}_\mathrm{ahp} = \frac{\mathbf{U}_\mathrm{hp}}{\mathbf{U}_\mathrm{da}}, \text{ and } \quad \mathbf{H}_{\mathrm{o}_\mathrm{hp}} = \frac{\mathbf{U}_\mathrm{hp}}{\mathbf{S}_\mathrm{hp}} = \mathbf{H}_\mathrm{da} \cdot \mathbf{H}_\mathrm{ahp}. \tag{2.8}$$

Measurements, recordings, and playback situations discussed in this thesis are implemented digitally at 44.1 kHz sample rate with custom-made software (Völk et al. 2007, 2009b). IRs are estimated using the exponential sine sweep (ESS) method according to Farina (2000) and Müller and Massarani (2001). The D/A- and A/D-converters employed[1] provide 24 Bit word length encoded by the interface driver in 32 Bit fixed point representation, while processing and storage are implemented at double precision (64 Bit floating point).

2.3 Critical Bandwidth and Critical-Band Rate

The concept of critical bands as introduced based on studies of Fletcher and Munson (1937) by Zwicker et al. (1957) describes frequency bands of frequency dependent spectral width with no fixed position on the frequency scale, as they appear in various psychoacoustic experiments (cf. Fastl and Zwicker 2007, pp. 150–158). Among those experiments are studies on loudness summation (Zwicker et al. 1957), absolute thresholds (Gässler 1954), and masking patterns of narrow-band noises (Fastl and Schorer 1986). It is the common characteristic of all these experiments to show considerably different results if the spectral width of one of the stimuli involved is increased beyond the critical bandwidth (CBW) $\Delta f_{\mathrm{G}}(f)$ centered at the respective frequency f.

Based on the critical bands, a critical-band rate (CBR) function $z(f)$ has been proposed (Zwicker 1961b), relating frequency to CBR so that all critical bands are equally wide on the CBR scale. Different formulae to calculate the CBR, its inverse $f(z)$, and the CBW have been proposed (Zwicker and Terhardt 1980, Traunmüller 1990), and the critical-band concept is frequently applied, for example in auditory-adapted short-term Fourier transform (Fourier-t transform, Terhardt 1985), speech coding (Mummert 1997), signal analysis (Völk et al. 2009b, 2011c), and signal processing for auditory prostheses such as hearing aids (Chalupper and Fastl 2002) and cochlear implants (Loizou 1998).

Previously introduced CBW functions specify the bandwidth spectrally symmetric around a center frequency, while not converging to zero at $f = 0\,\mathrm{Hz}$. This fact confines the applicability and universality of the concept, since a band-pass bandwidth $\Delta f(f) > 2f$ centered at f can hardly be justified by psychoacoustic experiments, even if ideal filtering is assumed, since the human hearing system produces almost no sensations in the frequency range below some 20 Hz (Fastl and Zwicker 2007, p. 18). Consequently, there is no reliable auditory incentive for a specific shape of the CBW function in this frequency range. The current expressions appear to be selected in a way to keep the formulae simple and the functions continuous, as supported by Fastl and Zwicker (2007, p. 158):

> *"Although the lowest critical bandwidth in the audible frequency region may be very close to 80 Hz, it is attractive to add the inaudible range from 0 Hz to 20 Hz to that critical band, and to assume that the lowest critical band ranges from 0 Hz to 100 Hz."*

However, it is impossible to implement bandwidths $\Delta f(f) > 2f$ with simple signal processing approaches or analysis systems specified in the positive frequency range. Using more elaborate signal processing algorithms with CBW filtering, the given low-frequency shape results in insufficient selectivity in the low-frequency range, producing artifacts that can be reduced by decreasing the low-frequency filter bandwidth (cf. Mummert 1997,

[1] RME Fireface 400, RME Fireface UC, and RME Multiface II, analog dynamic range $\Delta L_{\mathrm{D}} > 100\,\mathrm{dB}$

pp. 9–12). The analytic expression for the frequency dependence of the CBR proposed by Zwicker and Terhardt (1980) is based on values tabulated in the frequency range $0\,\mathrm{Hz} < f < 15.5\,\mathrm{kHz}$. At higher frequencies, this formula underestimates the CBR, as shown below. In addition, the function is not invertible in closed form, not allowing for computation of the frequency corresponding to a given CBR. However, this invertibility is crucial for algorithms aiming at equally distributed processing on the CBR scale.

In this section, a refined analytic expression for the CBW is proposed in dependence of $0\,\mathrm{Hz} \leq f \leq 20\,\mathrm{kHz}$, similar to the well-established expression, but by a continuous function with $\Delta f(f) \leq 2f\ \forall f$, while representing the values originally tabulated. Further, a refined and invertible CBR function is derived for the same frequency range. The formulae are defined based on an overview on earlier approaches and are especially intended to allow for direct parameterization of auditory-adapted signal processing routines based on the critical band concept, as for example the auditory-adapted analysis and the auditory-adapted exponential transfer function smoothing introduced in sections 2.4 and 5.1.5.

Critical Bandwidth: Concept and Earlier Formulae The CBW has been estimated as a function of frequency level independently based on psychoacoustic measurements with different methods on more than 50 subjects (Fastl and Zwicker 2007, p. 185). Originally, the CBWs $\Delta f_\mathrm{G}\,[n]$ were given by Zwicker (1961b) in table form as sample points $f_\mathrm{c}\,[n]$ with $n = 1, \ldots, 24$, indicated by the black dots in figure 2.2.

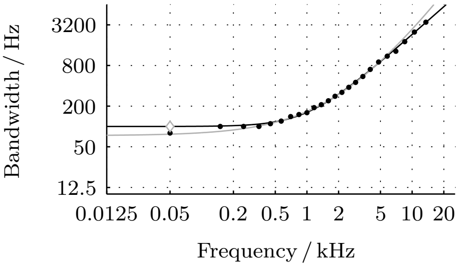

Figure 2.2: Critical bandwidth $\Delta f_\mathrm{G}\,(f)$ as a function of frequency f according to Zwicker and Terhardt (1980, black contour) and Traunmüller (1990, gray contour). The black dots indicate the originally tabulated values according to Zwicker (1961b), the unfilled diamond shows the updated value according to Zwicker and Terhardt (1980).

The tabulated values were reprinted by Zwicker and Terhardt (1980) with a modification: the lowest CBW $\Delta f_\mathrm{G}\,[1] = 80\,\mathrm{Hz}$ was changed to 100 Hz (diamond in figure 2.2, cf. Fastl and Zwicker 2007, p. 160), to get approximately constant CBWs for $n = 1, \ldots, 5$ that is at frequencies below about 500 Hz (Fastl and Zwicker 2007, p. 158). Based on the updated values, Zwicker and Terhardt (1980) proposed the frequency dependent CBW function

$$\frac{\Delta f_{\mathrm{G_Z}}\,(f)}{\mathrm{Hz}} = 25 + 75 \left[1 + 1.4 \left(\frac{f}{\mathrm{kHz}}\right)^2\right]^{0.69}, \tag{2.9}$$

fitting the updated data with an accuracy of ±10% (Fastl and Zwicker 2007, p. 164, black contour in figure 2.2), while deviating by 25% from the original value for $n = 1$.

Traunmüller (1990) derived a simplified set of analytic expressions for the application of the critical-band concept to speech technology. The CBW was defined by Traunmüller in the range $0.27\,\mathrm{kHz} < f < 5.8\,\mathrm{kHz}$ as a function of CBR (cf. below) by

$$\frac{\Delta f_{\mathrm{G_T}}\left(z_{\mathrm{T}}\left(f\right)\right)}{\mathrm{Hz}} = 52548\left[\left(\frac{z_{\mathrm{T}}\left(f\right)}{\mathrm{Bark}}\right)^2 - \frac{52.56 z_{\mathrm{T}}\left(f\right)}{\mathrm{Bark}} + 690.39\right]^{-1}. \tag{2.10}$$

Equation 2.10 is shown over the full audible frequency range by the gray contour in figure 2.2 since a function valid for the whole audio spectrum is targeted here.

Critical Bandwidth: Extensions Both previous formulae for the frequency dependence of the CBW do *not* exhibit the properties desired here. The most important requirement is little deviation from the original CBW function $\Delta f_{\mathrm{G_Z}}\left(f\right)$, given by equation 2.9. Further, $\Delta f_{\mathrm{G_V}}\left(0\,\mathrm{Hz}\right) = 0\,\mathrm{Hz}$ and $\Delta f_{\mathrm{G_V}}\left(f\right) \leq 2f\ \forall f$ are desired. The function

$$\Delta f_{\mathrm{G_V}}\left(f\right) = \Delta f_{\mathrm{G_Z}}\left(f\right)\left(1 - \frac{1}{\left(38.73 f/\,\mathrm{kHz}\right)^2 + 1}\right), \quad 0\,\mathrm{Hz} \leq f \leq 20\,\mathrm{kHz} \tag{2.11}$$

fulfills both requirements while fitting the sample values $\Delta f_{\mathrm{G}}\left[n\right]$ tabulated by Zwicker (1961b) with an accuracy of $\pm 10\%$ for $n = 1, \ldots, 24$. The black contour in figure 2.3 shows equation 2.11 along with the values tabulated by Zwicker (1961b, dots) and $\Delta f_{\mathrm{G_Z}}\left(f\right)$ proposed by Zwicker and Terhardt (1980, solid gray contour). The dashed gray lines indicate $f = 20\,\mathrm{Hz}$, approximately the lowest audible frequency, and $\Delta f(f) = 2f$.

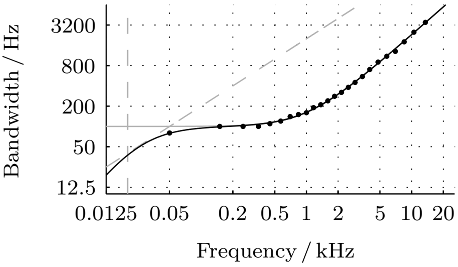

Figure 2.3: Critical bandwidth $\Delta f_{\mathrm{G}}\left(f\right)$ as a function of frequency f according to Zwicker and Terhardt (1980, solid gray contour) and to the equation proposed here (black contour). The black dots indicate the tabulated original values according to Zwicker (1961b). The dashed gray lines mark $\Delta f(f) = 2f$ and an approximation of the lower limit of the audible frequency range at $f = 20\,\mathrm{Hz}$.

Critical-Band Rate: Concept and Earlier Formulae According to Fastl and Zwicker (2007, p. 158), the CBR is developed based on the fact that the human hearing system analyzes broadband sounds in spectral sections corresponding to the critical bands. Consequently, the frequency dependence of the CBR, assigned the unit *Bark*, is helpful for understanding and modeling characteristics of the human hearing system. The CBR function $z\left(f\right)$ is according to Fastl and Zwicker (2007, p. 159) defined by an interpolation of the integer valued sample points $z\left[m\right] = m\,\mathrm{Bark}$, with $m = 0, \ldots, 24$, corresponding to the frequencies $f_{\mathrm{l}}\left[m\right]$ (black dots in figure 2.4, Fastl and Zwicker 2007, p. 160).

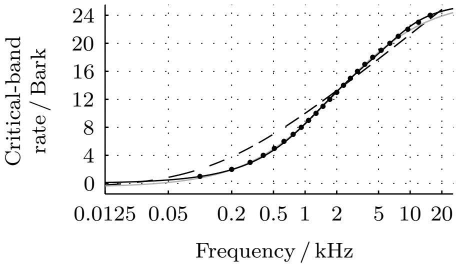

Figure 2.4: Critical-band rate $z(f)$ as a function of frequency f according to Zwicker and Terhardt (1980, black contour), Traunmüller (1990, gray contour), and Greenwood (1961, 1990, black dashed contour). The black dots indicate the originally tabulated values according to Zwicker (1961b).

The frequencies $f_\mathrm{l}[m]$ are given by the limits of 24 critical bands seamlessly arranged on the frequency scale, beginning at $f_\mathrm{l}[0] = 0\,\mathrm{Hz}$, so that the upper limiting frequency of each band with the center frequency $f_\mathrm{c}[\kappa]$ equals the lower limiting frequency of the next higher band according to

$$f_\mathrm{l}[\kappa+1] = f_\mathrm{l}[\kappa] + \Delta f_{\mathrm{G}_\mathrm{Z}}(f_\mathrm{c}[\kappa]), \quad \kappa = 0, \ldots, 23, \tag{2.12}$$

where $\Delta f_{\mathrm{G}_\mathrm{Z}}(f_\mathrm{c}[\kappa])$ grows with κ that is with frequency (Fastl and Zwicker 2007, p. 160, black dots in figure 2.2). Zwicker and Terhardt (1980) proposed, based on the sample values given by Zwicker et al. (1957) and Zwicker (1961b), the analytic expression

$$\frac{z_\mathrm{Z}(f)}{\mathrm{Bark}} = 13 \arctan \frac{0.76 f}{\mathrm{kHz}} + 3.5 \arctan \left(\frac{f}{7.5\,\mathrm{kHz}}\right)^2 \tag{2.13}$$

for the frequency dependence of the CBR (cf. also Fastl and Zwicker 2007, p. 164 and the solid black contour in figure 2.4). Equation 2.13 fits the original sample points with an accuracy of $\pm 0.2\,\mathrm{Bark}$.

However, the applicability of equation 2.13 for the parameterization of auditory-adapted algorithms is limited, since it is not invertible in closed form. Furthermore, $z_\mathrm{Z}(f)$ has been proposed based on values tabulated in the frequency range $0\,\mathrm{Hz} \leq f \leq 15.5\,\mathrm{kHz}$, while nowadays hardware and algorithms are often designed to process signals with bandwidths exceeding the audible frequency range $20\,\mathrm{Hz} \leq f \leq 20\,\mathrm{kHz}$.

Unfortunately, $z_\mathrm{Z}(f)$ tends to underestimate the CBR at frequencies $f > 16\,\mathrm{kHz}$. For example, $z_\mathrm{Z}(20\,\mathrm{kHz}) \approx 24.58\,\mathrm{Bark}$ holds, while the CBW of the highest critical band tabulated is given to $\Delta f_{\mathrm{G}_\mathrm{Z}}(f_\mathrm{l}[23]) = 3.5\,\mathrm{kHz}$. If the CBW is assumed to continue growing disproportionately with frequency for $f > 15.5\,\mathrm{kHz}$, a bandwidth in the range of $\Delta f_{\mathrm{G}_\mathrm{Z}}(f_\mathrm{l}[24]) = 4.5\,\mathrm{kHz}$ appears a reasonable estimate. Hence, according to equation 2.12,

$$f_\mathrm{l}[25] = f_\mathrm{l}[24] + \Delta f_{\mathrm{G}_\mathrm{Z}}(f_\mathrm{l}[24]) = (15.5 + 4.5)\,\mathrm{kHz} = 20\,\mathrm{kHz} \tag{2.14}$$

holds true and the hypothetic CBR sample value $z[25] = 25\,\mathrm{Bark}$ would be reached at $f_\mathrm{l}[25] = 20\,\mathrm{kHz}$. Consequently, $z_\mathrm{Z}(20\,\mathrm{kHz}) \approx 24.58\,\mathrm{Bark}$ computed using equation 2.13 represents an underestimate of the CBR at $f = 20\,\mathrm{kHz}$.

The analytic expressions proposed by Traunmüller (1990) contain an invertible function describing the dependence of the CBR on frequency by

$$\frac{z_{\mathrm{T}}\left(f\right)}{\mathrm{Bark}} = 26.81\,\frac{f/\,\mathrm{Hz}}{1960 + f/\,\mathrm{Hz}} - 0.53. \tag{2.15}$$

This function is defined in the frequency range $200\,\mathrm{Hz} < f < 6.7\,\mathrm{kHz}$, where it fits the original samples with an accuracy of $\pm 0.05\,\mathrm{Bark}$, while deviating at frequencies outside this range by up to $0.73\,\mathrm{Bark}$ (gray contour in figure 2.4). Especially, $z_{\mathrm{T}}\left(0\,\mathrm{Hz}\right) \neq 0\,\mathrm{Bark}$ holds true, which is not in accordance with the definition of the CBR.

Assuming that the positions on the basilar membrane in the inner ear correspond to the CBR, Greenwood (1961, 1990) derived relations between CBR and frequency from a cochlear frequency-position function based on physiological data of different species. The function relating frequency to CBR is given for human listeners (Greenwood 1990, dashed black contour in figure 2.4) by

$$\frac{z_{\mathrm{G}}\left(f\right)}{\mathrm{Bark}} = 11.9\,\log_{10}\left(\frac{f}{165.4\,\mathrm{Hz}} + 0.88\right). \tag{2.16}$$

While invertible, equation 2.16 deviates up to $2.17\,\mathrm{Bark}$ from the sample points originally tabulated by Zwicker (1961b).

Critical-Band Rate: Extensions None of the formulae introduced previously fulfills the requirements requested here. Therefore, the invertible relation of CBR to frequency

$$\frac{z_{\mathrm{V}}\left(f\right)}{\mathrm{Bark}} = 32.12\left\{1 - \left[1 + \left(\frac{f/\,\mathrm{Hz}}{873.47}\right)^{1.18}\right]^{-0.4}\right\}, \quad 0\,\mathrm{Hz} \leq f \leq 20\,\mathrm{kHz} \tag{2.17}$$

is proposed, approximating the values tabulated by Zwicker (1961b) with an accuracy of $\pm 0.08\,\mathrm{Bark}$, in contrast to $\pm 0.2\,\mathrm{Bark}$ achieved by the noninvertible equation 2.13. Furthermore, equation 2.17 is fitted to the tabulated values extended by the pair $f_{\mathrm{l}}\left[25\right] = 20\,\mathrm{kHz}$ and $z\left[25\right] = 25\,\mathrm{Bark}$. For that reason, evaluating equation 2.17 at $f = 20\,\mathrm{kHz}$ results in the realistic value $z_{\mathrm{V}}\left(20\,\mathrm{kHz}\right) \approx 24.86\,\mathrm{Bark}$.

The inverse of equation 2.17, required if frequencies corresponding to a pre-defined distribution (e. g. equally spaced) on the CBR scale are to be computed, is given by

$$\frac{f_{\mathrm{V}}\left(z\right)}{\mathrm{Hz}} = 873.47\left[\left(\frac{32.12}{32.12 - z/\,\mathrm{Bark}}\right)^{2.5} - 1\right]^{1/1.18}, \quad 0\,\mathrm{Bark} \leq z \leq 24.86\,\mathrm{Bark}. \tag{2.18}$$

In combination, the proposed formulae allow for the parameterization of auditory-adapted signal processing routines directly, which is used in the following section to derive an auditory-adapted system analysis procedure based on the critical-band concept.

2.4 Auditory-Adapted Analysis

The IRs of LTI systems represent a simple but complete system description. For that reason, IR based system analysis is attempted here, wherever possible even for time-variant systems. Mathematically, processing an audio signal by a specific digital or analog LTI system is described in that the signal, without loss of generality assumed digitized according to section 2.2, is convolved with the system IR, represented by a finite impulse response (FIR) approximation here. This process represents filtering the audio signal with the filter described by the FIR. That way, the temporal and spectral characteristics of the filter are superimposed on the signal finally detected by the hearing system.

When dealing with LTI systems, analyzing one IR is sufficient to describe the system (Oppenheim et al. 1998), while piecewise time-invariant systems can be approximated by series of IRs (cf. section 5.1.1). For that reason, it is possible to predict the audible impact of a system approximately by analyzing an FIR or a series of FIR approximations in a way resembling the time-frequency analysis of the hearing system.

Auditory Adaption in General In the following, an auditory-adapted system analysis method is derived, aiming at visually representing perceptually relevant spectro-temporal system characteristics based on an FIR approximation of the system IR. Since the hearing system represents a complex adaptive and nonlinear system (Fastl and Zwicker 2007, p. 50), auditory adaption is aimed at regarding aspects considered primarily relevant in the context of virtual acoustics. Following Terhardt (1998, p. 223) and taking into account the findings of section 5.4, the sound pressure signals at the eardrums may be considered the primary inputs to the human hearing system. According to Terhardt (1998, p. 230), the subsequent stage of the hearing system, the level dependent and nonlinear middle ears, may be approximated by TFs from the eardrum pressures to the oval window velocities, the latter in turn considered the input signals to the inner ears.

The signal analysis of the human inner ear depends on the signal level, spectrum, and temporal shape (Fastl 1982, Zwicker 1984, 1986, Fastl and Zwicker 2007). It is this complex nature of the analysis (essentially influenced by the active and nonlinear cochlea, Zwicker 1979, 1985) that allows normal-hearing listeners to communicate even under adverse conditions (Hojan and Fastl 1996, Stemplinger et al. 1997, Rader et al. 2008). At the same time, however, it aggravates auditory-adapted signal processing, especially for signal analysis (e. g. loudness prediction or noise assessment, Zwicker 1977, Fastl 2000, Fastl et al. 2009), rehabilitation of hearing disorders with reduced spectral or temporal selectivity (Zwicker and Bubel 1977, Fastl et al. 1998, Chalupper and Fastl 2002, Qin and Oxenham 2003), and audio playback in noisy environments (due to partial masking, Gleiss and Zwicker 1964, Zwicker 1987, 1989).

For the purpose of system characterization, a level and signal independent analysis is desirable. In contrast to the auditory-adapted *signal* analysis, as for example by Fourier-t transform (Terhardt 1985) or filter-bank approaches (as frequently used in loudness prediction, cf. Zwicker et al. 1984), *system* analysis is not intended specifically for one signal; the results are to be representative for all signals possibly processed by the system under consideration. For that reason, signal dependent aspects must be studied for the specific signal and cannot be included in a general system analysis. With respect to the hearing system, this means simplification of the analysis procedure.

Modeling Aspects A spectrum analysis system resolves two spectral lines separated in frequency more than the analysis bandwidth. According to Zollner and Zwicker (1993, pp. 286–287), the spectrum analysis of the human inner ear can be modeled by a number of overlapping band-pass filters with frequency dependent so-called critical bandwidths (CBWs, cf. section 2.3). Based on this filter-bank model, the frequency dependent minimum temporal resolution of the hearing system's spectrum analysis is inversely proportional to the CBW (T. Horn, personal communication, August 2011). Following definition 3, this resolution is referred to as monaural temporal resolution.

Definition 3 (*Monaural and Binaural Hearing, Attributes, and Resolution*)

Hearing with two ears is referred to as binaural, in contrast to monaural hearing with one ear. Binaural attributes and localization cues occur only in binaural hearing, all other attributes and cues are monaural. The frequency dependent temporal resolution of the hearing system's monaural spectral analysis is referred to as monaural in contrast to the binaural temporal resolution of interaural differences.

The CBW grows according to equation 2.11 with frequency, ranging from values around $\Delta f_{\mathrm{G}}\,(20\,\mathrm{Hz}) \approx 80\,\mathrm{Hz}$ at low frequencies to $\Delta f_{\mathrm{G}}\,(20\,\mathrm{kHz}) \approx 6\,\mathrm{kHz}$ in the highest audible frequency range (section 2.3). Assuming for the hearing system, in accordance to technical systems (cf. Kammeyer 1992, p. 50), a constant bandwidth-period product

$$\Delta f_{\mathrm{G}}\,(f)\,T_{\mathrm{G}}\,(f) = 1, \tag{2.19}$$

the CBWs correspond to frequency dependent periods between $T_{\mathrm{G}}\,(20\,\mathrm{Hz}) \approx 10\,\mathrm{ms}$ and $T_{\mathrm{G}}\,(20\,\mathrm{kHz}) \approx 160\,\mu\mathrm{s}$. The period $T_{\mathrm{G}}\,(f)$ corresponds well to psychoacoustic results on the monaural temporal resolution, especially at frequencies below about 1 kHz (Wiegrebe et al. 1996, Krumbholz et al. 2003). Masking experiments largely confirm the monaural temporal resolution model applied here at $f = 300\,\mathrm{Hz}$ and $f = 900\,\mathrm{Hz}$, while indicating considerably less temporal selectivity than predicted by the monaural temporal resolution at $f = 2.7\,\mathrm{kHz}$ and $f = 8.1\,\mathrm{kHz}$ (Plack and Moore 1990), probably influenced by the subsequent processing by the hearing system not taken into account here. This being said, auditory-adapted system analysis must consider temporal aspects of systems with IRs containing auditory relevant energy extended beyond the monaural temporal resolution. Such systems are after definition 4 referred to as auditory spectro-temporally effective here, in contrast to purely spectrally effective systems.

Definition 4 (*Spectro-Temporally and Spectrally Effective Systems*)

The critical-band wide band-pass filtered impulse response of a spectro-temporally effective system contains auditory relevant energy extended in time beyond the filter slope corrected monaural temporal resolution. Other systems are purely spectrally effective. Auditory relevant impulse response amplitudes lie less than the dynamic range of the hearing system below the impulse response maximum.

Since resolution and IRs may depend on frequency, the comparison must take place in each channel of the critical-band wide band-pass filtered IR. Temporal aspects are to be considered if auditory relevant energy extends in at least one channel beyond the respective monaural temporal resolution corrected by the temporal band-pass filter slope. Auditory relevant IR amplitudes are assumed to lie in the present context less than the dynamic range of the hearing system below the maximum IR amplitude. In the frequency domain, the TFs of spectro-temporally effective systems show suprathreshold spectral variation within the critical bands containing temporally effective sections.

Typical room reverberation times exceed 100 ms (Schroeder et al. 1966, Kuttruff and Jusofie 1967). For that reason, most reverberant rooms represent spectro-temporally effective systems. Loudspeaker IRs show decay times between about 1 ms and 30 ms, depending on the design concept (Fincham 1985, Adams 1989, Hawksford 1997a, Dyreby and Choisel 2007), and may therefore be spectrally or spectro-temporally effective. Most other audio processing systems, as for example headphones, amplifiers, or audio interfaces exhibit decay times in the range of the monaural temporal resolution (Kuttruff and Jusofie 1969, Hirahara 2004). In general, it has to be proven by measurement whether a specific system is effective spectro-temporally or purely spectrally.

In this thesis, the characteristics of purely spectrally effective systems are analyzed visually by auditory-adapted frequency dependent TFs, whereas for spectro-temporally effective systems, auditory-adapted time and frequency dependent spectrogram representations are shown. When analyzing separately two systems employed to process corresponding signals designated for the left and right ears, disparities have to be addressed in addition to the single system characteristics, because the hearing system analyzes interaural differences (Mills 1972, Middlebrooks and Green 1991, Blauert 1997). The binaural temporal resolution according to definition 3 reaches values in the range of 10 μs, exceeding the monaural temporal resolution (Hershkowitz and Durlach 1969, Domnitz 1973, Fastl and Zwicker 2007, p. 293). Two systems can be compared by time and frequency dependent auditory-adapted interaural spectrograms, justified by the independence of the interaural level difference thresholds for broadband noise on interaural correlation, suggesting independent processing of both ear signals prior to interaural comparison processes (Hartmann and Constan 2002).

Implementation and Parameterization of the Auditory-Adapted Analysis In 1985, Terhardt proposed a method for auditory-adapted audio *signal* analysis referred to as Fourier-t transform (FTT), which is extended here to the application for auditory-adapted *system* analysis. As the resulting procedure is capable of analyzing systems and signals, it is generally referred to as auditory-adapted analysis (AAA). Based on equation 3.74 of Terhardt (1998), the FTT spectrogram

$$s_\kappa\,[n] = \frac{1}{f_\mathrm{s}} \sum_{i=0}^{n} s\,[i]\, w_\kappa\,[n-i]\ \mathrm{e}^{-\,\mathrm{j}2\pi f_{\mathrm{A}_\kappa} i/f_\mathrm{s}}, \quad \kappa = 0,\,\ldots,\,\Upsilon - 1 \tag{2.20}$$

of the causal sequence $s\,[n]$ sampled at f_s with $n = 0,\,\ldots,\,N-1$ is computed at the analysis frequencies f_{A_κ} using a series of causal window functions $w_\kappa\,[n]$ allowing adapting the transformation to the time-frequency resolution of the hearing system. According to Terhardt (1985, p. 253 and Mummert 1997, p. 9), the ascending exponential functions

$$w_\kappa\,[n] = \begin{cases} 2a_\kappa\,\mathrm{e}^{-a_\kappa n/f_\mathrm{s}}, & n \geq 0, \\ 0, & n < 0 \end{cases} \tag{2.21}$$

represent appropriate temporal window functions. Following Terhardt (1985, equation 38), each $w_\kappa\,[n]$ represents a first-order low-pass filter with the 3 dB-bandwidth $\Delta f_\kappa = a_\kappa/\pi$. Schlang and Mummert (1990) extend the special case of equation 2.21 to the more general class of exponential temporal windows of arbitrary integer order $\eta = 1, 2, \ldots$ with η-fold pole at a_κ. This series of causal window functions is given by

$$w_{\eta_\kappa}\,[n] = \begin{cases} \dfrac{2a_\kappa}{(\eta-1)!}\,(a_\kappa n/f_\mathrm{s})^{\eta-1}\,\mathrm{e}^{-a_\kappa n/f_\mathrm{s}}, & n \geq 0, \\ 0, & n < 0, \end{cases} \tag{2.22}$$

with the corresponding 3 dB-bandwidths

$$\Delta f_{\eta_\kappa} = (a_\kappa/\pi)\,\sqrt{2^{1/\eta} - 1}. \tag{2.23}$$

According to Mummert (1997, pp. 90–91) $\eta = 4$ represents a typical value for auditory system modeling also optimal for re-synthesis (Patterson et al. 1992, Irino and Patterson 2001). Therefore, $\eta = 4$ is selected here. For $\eta = 1$, equation 2.22 equals equation 2.21. In this case, all the window functions reach their maximum at $n_{\mathrm{max},1} = 0$. For $\eta > 1$, the window maxima occur delayed by $n_{\mathrm{max},\eta_\kappa} > 0$. The sample indices corresponding to the maximum amplitudes of the window functions are derived in appendix D.2 to

$$n_{\mathrm{max},\eta_\kappa} = (\eta - 1)\,f_\mathrm{s}/a_\kappa, \quad \forall\,\eta > 1. \tag{2.24}$$

The delay of the maximum window amplitude decays with increasing frequency and depends on the order η. For AAA, the delay is desired to reflect the frequency dependent travel times to the positions of maximum excitation on the basilar membrane in the human inner ear, given by table 2.1 according to von Békésy (1949).

f_res/ kHz	0.05	0.1	0.2	0.5	0.7	2
τ_res/ ms	6	3	2	1	0.7	0.2

Table 2.1: Approximate travel times τ_res of a pulse wave along the basilar membrane in the human cochlea to the positions of maximum excitation after von Békésy (1949), expressed by the corresponding resonant frequency f_res.

The travel times indicate that the spectral components of a Dirac impulse are not processed simultaneously in the cochlea, but that the detection of low-frequency components is delayed compared to higher frequencies. This fact may be reflected qualitatively in AAA with windows according to equation 2.22 by adjusting the delays of the window maxima using equation 2.24 so that

$$n_{\mathrm{max},\eta_\kappa} \approx \tau_{\mathrm{res}_\kappa} f_{\mathrm{s}} \tag{2.25}$$

holds true. Since $\eta = 4$ is desired and f_{s} is constant, the parameter a_κ controls the degree to which equation 2.25 is fulfilled. Based on equation 2.23, the frequency dependent analysis bandwidth is adapted to given bandwidths Δf_{η_κ} by selecting

$$a_\kappa = \pi \left(\Delta f_{\eta_\kappa} / \sqrt{2^{1/\eta} - 1} \right). \tag{2.26}$$

For AAA, it is desirable to select the analysis bandwidth proportional to the analysis bandwidth of the hearing system given by the CBW (equation 2.11). Introducing the constant multiplication factor c, the CBW proportional 3 dB-bandwidth is given by

$$\Delta f_{\eta_\kappa} = c \, \Delta f_{\mathrm{G_V}} \left(f_{\mathrm{A}_\kappa} \right). \tag{2.27}$$

Selecting $c = 0.5$ according to Mummert (1997, pp. 91–93) fulfills equation 2.25 in good approximation and is also supported by functional inner ear simulations of Peisl (1990). Consequently, this value is chosen for AAA here. The CBW definition in section 2.3 is derived with the aim of suppressing undesired selection of spectral contributions at negative frequencies, especially for low analysis frequencies. Depending on the sampling frequency, alias artifacts may occur at the upper limit of the analysis frequency range due to the finite steepness of the analysis filter slopes. Therefore, the sampling frequency represents a critical parameter. In the course of this thesis, $f_{\mathrm{s}} = 44.1\,\mathrm{kHz}$ has proven useful. Following Terhardt (1985), the analysis is carried out at the $\Upsilon = 625$ frequencies

$$f_{\mathrm{A}_\kappa} = f_{\mathrm{V}} \left(z_{\mathrm{V}} \left(20\,\mathrm{Hz} \right) + \frac{z_{\mathrm{V}} \left(20\,\mathrm{kHz} \right) - z_{\mathrm{V}} \left(20\,\mathrm{Hz} \right)}{\Upsilon - 1} \kappa \right), \quad \kappa = 0, 1, \ldots, \Upsilon - 1, \tag{2.28}$$

equally spaced on the CBR scale between 20 Hz and 20 kHz (using equations 2.17 and 2.18). This results in a resolution equal to the just noticeable change of frequency (Fastl and Zwicker 2007, pp. 182–187), reflected in about 0.04 Bark analysis step size in the CBR range between approximately 0.15 and 24.86 Bark. According to Mummert (1997, p. 93), resolutions below 0.05 Bark are also sufficient to avoid visualization artifacts.

Analysis Procedure and Visualization The AAA process as proposed here starts for a specific linear acoustic system by acquiring an FIR approximation of the system IR. For time-invariant systems, one FIR is sufficient, while for time-variant systems a set of FIRs covering all system states of interest is analyzed. The following steps are described exemplary for an LTI system (audio interface[2] in short-circuit state between in- and output) and therefore for one IR approximated by the FIR $h_{\mathrm{au,sc}}[n]$. AAA is carried out by setting $s[n] = h_{\mathrm{au,sc}}[n]$ in equation 2.20, resulting in general in the auditory-adapted

[2] RME Fireface UC, $f_s = 44.1\,\mathrm{kHz}$, 256 samples buffer, balanced line short-circuit, 13 dBu output level

spectrogram $h_{\mathrm{AAA}_\kappa}[n]$ and resulting here in the audio interface spectrogram $h_{\mathrm{au,sc}_\kappa}[n]$. With regard to the nomenclature of equation 2.20, this is formulated by

$$s_\kappa[n] = h_{\mathrm{AAA}_\kappa}[n] = h_{\mathrm{au,sc}_\kappa}[n]. \tag{2.29}$$

In conventional auditory-adapted signal analysis, $s_\kappa[n]$ is visualized by two separate gray-scale bitmap images, the auditory magnitude and phase spectrograms, with the abscissae representing the temporal and the ordinates the spectral dimension (Horn 1998, Mummert 1997, p. 134). The results are usually displayed over the linear CBR scale, with the axis labeled by the corresponding frequencies. In other words, the frequency scale is warped in an auditory-adapted manner. According to Mummert (1997, equation 1.10), each pixel corresponds to a complex-valued matrix element of the band-pass spectrogram

$$h^{\mathrm{bp}}_{\mathrm{AAA}_\kappa}[n] = h_{\mathrm{AAA}_\kappa}[n]\, \mathrm{e}^{\mathrm{j}2\pi f_{\mathrm{A}_\kappa} n/f_\mathrm{s}}. \tag{2.30}$$

The level of gray represents in the auditory magnitude spectrogram the element's logarithmic magnitude

$$L_\kappa[n] = 20\log_{10}\left|h_{\mathrm{AAA}_\kappa}[n]\right|\mathrm{dB} = 20\log_{10}\left|h^{\mathrm{bp}}_{\mathrm{AAA}_\kappa}[n]\right|\mathrm{dB} \tag{2.31}$$

and in the auditory phase spectrogram the phase angle

$$A_\kappa[n] = \begin{cases} \arg\left(h^{\mathrm{bp}}_{\mathrm{AAA}_\kappa}[n]\right) & \text{if } \arg\left(h^{\mathrm{bp}}_{\mathrm{AAA}_\kappa}[n]\right) \geq 0, \\ \arg\left(h^{\mathrm{bp}}_{\mathrm{AAA}_\kappa}[n]\right) + \pi & \text{otherwise.} \end{cases} \tag{2.32}$$

Figure 2.5 shows the auditory magnitude and phase spectrograms of $h_{\mathrm{au,sc}}[n]$. The black dots in the magnitude spectrogram (left diagram) indicate the frequency dependent travel times $\tau_{\mathrm{res}_\kappa}$ along the basilar membrane according to von Békésy (1949, cf. table 2.1) with respect to the maximum amplitude of the broadband IR.

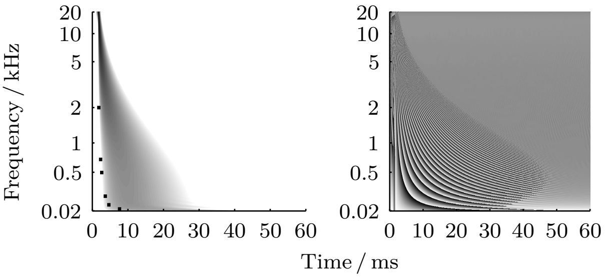

Figure 2.5: Auditory-adapted logarithmic magnitude spectrogram with 60 dB visible dynamic (left) and phase spectrogram (0 to 2π, right) of an audio interface impulse response in short-circuit state, approximated by a finite impulse response. The black dots in the left panel indicate travel times on the basilar membrane.

A visualization method referred to as AAA spectrogram is introduced here, combining the information of figure 2.5 in one image. The method is based on signal modification approaches by Horn, proven by informal listening to allow for transparent re-synthesis

(T. Horn, personal communication, August 2011). The incentive for the AAA spectrogram

$$L_{\mathrm{AAA}_\kappa}[n] = 20\log_{10}\Big(\big|h^{\mathrm{bp}}_{\mathrm{AAA}_\kappa}[n]\big|\Big\{1+\cos\big[\arg\big(h^{\mathrm{bp}}_{\mathrm{AAA}_\kappa}[n]\big)\big]\Big\}\Big)\,\mathrm{dB} \tag{2.33}$$

is in addition to condensed visualization by a single image the fact that the hearing system produces one analysis output at a time (time-frequency dependent excitation pattern, cf. Zwicker 1958, Fastl 1982), not a magnitude and a phase result. Figure 2.6 shows as an example the AAA spectrogram of the FIR approximation of the audio-interface impulse response $h_{\mathrm{au,sc}}[n]$ also represented by figure 2.5.

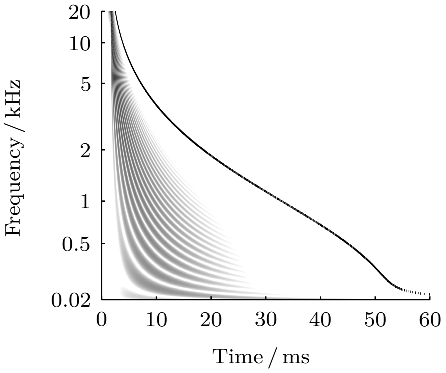

Figure 2.6: Auditory-adapted analysis spectrogram with 60 dB visible dynamic of an audio interface impulse response in short-circuit state, approximated by a finite impulse response. The black contour indicates the monaural temporal resolution given by the periods corresponding to the critical bandwidths, corrected by the filter slopes with respect to the impulse response maximum. Since the spectrogram decays by more than 60 dB within the monaural temporal resolution, the system is purely spectrally effective.

The basic idea of the AAA spectrogram is that the phase spectrogram can be effective only where the magnitude spectrogram shows suprathreshold levels. It appears reasonable to weigh the phase by the magnitude values, keeping magnitude information visible while showing relevant phase information. Direct weighting with the phase computed according to equation 2.32 causes discontinuities at the phase values 0 and 2π. For that reason, the phase cosine is included in equation 2.33, requiring the offset by one before taking the logarithm. Apart from this offset, the argument of the logarithm represents the real-part

$$\Re\big\{h^{\mathrm{bp}}_{\mathrm{AAA}_\kappa}[n]\big\} = \big|h^{\mathrm{bp}}_{\mathrm{AAA}_\kappa}[n]\big|\cos\big[\arg\big(h^{\mathrm{bp}}_{\mathrm{AAA}_\kappa}[n]\big)\big] \tag{2.34}$$

of the band-pass spectrogram $h^{\mathrm{bp}}_{\mathrm{AAA}_\kappa}[n]$ (Bronstein et al. 2001, equation 1.133a). The black contour in figure 2.6 indicates the periods corresponding to the CBWs with respect to the IR maximum, corrected by the 60 dB decay times of the analysis windows (cf. appendix D.2). This contour may be regarded as an approximation of the monaural temporal resolution. The audio interface AAA spectrogram decays at all audible frequencies by more than 60 dB within the monaural temporal resolution and is considered purely spectrally effective. Following the reverberation time definition (Schroeder 1965, DIN EN ISO 3382 2000), systems with AAA spectrograms decaying by at least 60 dB within the monaural temporal resolution are classified purely spectrally effective. This criterion is evaluated strictly at frequencies below 1 kHz here, where masking experiments confirm the procedure (Plack and Moore 1990). While the AAA spectrogram is used for the visual analysis of spectro-temporally effective systems, it is not necessary to study temporal aspects of spectrally effective systems, which are visualized based on the AAA spectrum

$$\bar{H}_{\mathrm{AAA}_\kappa} = \sum_{n=n_\mathrm{s}}^{N-1} h_{\mathrm{AAA}_\kappa}[n], \quad 0 \le n_\mathrm{s} < N-2 \tag{2.35}$$

by means of the logarithmic AAA magnitude spectrum

$$l_\kappa = 20\log_{10}\left|\bar{H}_{\mathrm{AAA}_\kappa}\right| \mathrm{dB} \tag{2.36}$$

and the AAA group delay

$$\tau_{\mathrm{g}_\kappa} = -\frac{\arg\left(\bar{H}_{\mathrm{AAA}_{\kappa+1}}\right) - \arg\left(\bar{H}_{\mathrm{AAA}_\kappa}\right)}{2\pi\left(f_{\mathrm{A}_{\kappa+1}} - f_{\mathrm{A}_\kappa}\right)}, \quad \kappa = 0, 1, \ldots, \Upsilon - 2. \tag{2.37}$$

In order to prevent phase ambiguities, the summation starting index n_s in equation 2.35 is selected so that no auditory relevant energy is present in $h_{\mathrm{AAA}_\kappa}[n]$ at sample indices

$$n \ge n_\mathrm{s} + \frac{f_\mathrm{s}}{2\left(f_{\mathrm{A}_{\kappa+1}} - f_{\mathrm{A}_\kappa}\right)} \tag{2.38}$$

for all analysis frequencies, that is for all κ (cf. equation 2.20). This is especially relevant for systems showing IRs with initial pure delay, where equation 2.38 can be typically fulfilled by selecting $n_\mathrm{s} > 0$ within the initial delay. The index computed by equation 2.38 depends on the analysis frequency spacing $\left(f_{\mathrm{A}_{\kappa+1}} - f_{\mathrm{A}_\kappa}\right)$ and therefore, with the parameterization employed here, on the number of analysis channels Υ (cf. equation 2.28). While $\Upsilon = 625$ has proven useful, higher values may be necessary for specific systems.

The magnitude spectrum and group delay according to equations 2.36 and 2.37 are referred to as AAA transfer characteristics. Figure 2.7 shows the AAA transfer characteristics of the audio interface FIR approximation depicted in figures 2.5 and 2.6.

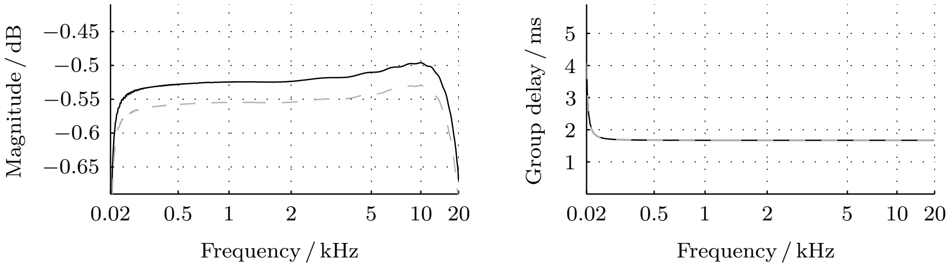

Figure 2.7: Auditory-adapted audio interface magnitude transfer function and group delay in short-circuit state of the analog in- and outputs. Balanced line connection, 13 dBu output level, cable lengths 1.5 m (solid black) and 46.5 m (dashed gray).

Since the AAA transfer characteristics are computed at analysis frequencies equally distributed on the CBR scale, they are shown over the linear CBR, the axes labeled with the corresponding frequencies. The abscissa represents a frequency scale warped in an auditory-adapted manner. The FIR $h_{\mathrm{au,sc}}$ is measured in balanced line short-circuit state of the analog in- and output channels between the corresponding digital sequences. The TF $H_{\mathrm{au,sc}} = H_{\mathrm{da}} \cdot H_{\mathrm{c}} \cdot H_{\mathrm{ad}}$ contains the D/A-converter TF H_{da} (equation 2.7), the A/D-converter TF H_{ad} (equation 2.5), and the cable TF H_{c}. Figure 2.7 shows results from two measurements, one with 1.5 m cable length ($H_{\mathrm{c,1.5\,m}}$, solid black) and one with 46.5 m cable length ($H_{\mathrm{c,46.5\,m}}$, dashed gray). Adding 45 m cable results in less than 0.04 dB AAA magnitude and less than 4.6 μs group delay difference. Considering the just noticeable differences in the ranges of 0.1 dB (Fastl and Zwicker 2007, p. 180) and 50 μs (Fastl and Zwicker 2007, p. 293) and assuming cable lengths of some 50 m to cover the scenarios discussed in this thesis, the influence of line connections is omitted. It might be desirable to consider connection TFs, which is covered by the framework introduced.

Variability in Auditory-Adapted Transfer Characteristics The variability in a set S of transfer characteristics is addressed using the quartiles, the 25 %, 50 %, and 75 % percentiles $\mathrm{P}_{25}(S)$, $\mathrm{P}_{50}(S)$, and $\mathrm{P}_{75}(S)$, computed according to equations 2.36 and 2.37 at each analysis frequency f_{A_κ} separately for l_κ and τ_{g_κ}. The auditory-adapted variability is defined comparable to the analysis of listening experiment results by the inter-quartile ranges of l_κ and τ_{g_κ} (cf. section 3.2.1), whereas other authors prefer arithmetic mean and standard deviation of the magnitude spectrum (e. g. Møller et al. 1995a). In order to keep possibly occurring asymmetries in the distribution visible, the partial variability values $\mathrm{V}_{25}(S) = \mathrm{P}_{25}(S) - \mathrm{P}_{50}(S)$ and $\mathrm{V}_{75}(S) = \mathrm{P}_{75}(S) - \mathrm{P}_{50}(S)$ are shown. If more than one set is discussed, the quartiles of the partial variability values of all data sets are given, referred to as the auditory-adapted variability statistics, and denoted for example in the case of $\mathrm{V}_{25}(S)$ by $\mathrm{P}_{25}(\mathrm{V}_{25}(S))$, $\mathrm{P}_{50}(\mathrm{V}_{25}(S))$, and $\mathrm{P}_{75}(\mathrm{V}_{25}(S))$.

2.5 Loudness Transfer Functions

A usual and feasible method to address the transmission characteristics of an electroacoustic system perceptively is the loudness comparison to the reference scene. The frequency dependent correction level necessary for narrow-band signals, presented by the system, to elicit the reference scene loudness is referred to as loudness transfer function (LTF).

Definition 5 (*Loudness Transfer Function*)

> *The frequency dependent correction levels necessary for an electroacoustic transmission system to elicit the reference scene loudness of narrow-band signals are referred to as loudness transfer function.*

Prominent LTFs are the perceptually measured free-field or diffuse-field HP equalization curves often employed for psychoacoustic experiments or HP reproduction in general (section 3.2.4, Zwicker and Maiwald 1963, Theile 1981).

Conceptual Aspects The loudness of an isolated fixed-frequency sinusoid depends on its level (Fletcher and Munson 1933, Robinson 1953, Robinson and Dadson 1956) and temporal shape (Fletcher and Munson 1933, Zwicker 1956, 1965). The loudness of more complex isolated steady stimuli depends in addition on the spectral content (Fletcher and Steinberg 1924, Steinberg 1925, Zwicker et al. 1957). For that reason, an LTF must not be confused with or misinterpreted as a system-theoretically defined magnitude spectrum. LTI systems are characterized fully by the signal independent IRs respectively TFs between their input and output ports (Oppenheim et al. 1998). It is possible to acquire the TF in a specific frequency range relating the spectrum of the system response to a known signal providing sufficient intensity in the frequency range of interest to the spectrum of that signal. A crucial prerequisite for the validity of this procedure is using an LTI device with frequency independent TF for recording the system response. If in contrast a frequency, level, or in general signal dependent measurement method as for example the human loudness perception is employed, a general TF computation by relating system output and input is not possible, because the system response recording may vary, due to the measurement method, with the specific signal.

LTFs are acquired by adjusting the same loudness that is the same *measurement system reading* in two situations: listening to the reference scene and listening to the device under test. Consequently, measurement system readings are compared, influenced by the nonlinear and signal dependent transfer characteristics of the measurement system, which is in the present case the hearing system. In comparing the readings of a measurement device with signal dependent transfer characteristics in response to the output signals of two different LTI devices under test, it must be taken into account that both measurement system readings are valid only for the specific signals at the *measurement device* input. It is not clear whether the specific measurement system input is the only signal resulting in the current reading, or if the reading would change for example with the signal duration. The only procedure allowing for the signal independent comparison of two LTI systems is comparing the system input signals leading to equal output signals or vice versa, necessarily requiring an LTI measurement system with frequency independent TF.

That being said, measuring the TFs of electroacoustic systems by the loudness they elicit is valid only if the measurement device inputs, the sound pressure time functions at the eardrums, are identical in the situations to be compared (according to Fastl and Zwicker 2007, p. 25, the ear is assumed pressure sensitive). Consequently, it is not possible in a system-theoretic sense to measure *“the frequency response of earphones* [. . .] *by subjectively performed loudness comparisons of tones, presented via a loudspeaker or via earphones”* (Fastl and Zwicker 2007, p. 8). This becomes even more evident considering the possibility of *different* sound pressure time functions at the eardrums eliciting the *same* loudness with loudspeaker and conventional headphone playback (cf. section 5.4 Fastl 1986, Völk et al. 2011d). In that case, the LTF of the HPs regarding *a specific* reference scene is measured, not the frequency response in general.

Procedure and Parameterization The experimental method proposed for LTF measurements is a loudness adjustment procedure using Békésy-Tracking according to Fastl and Zwicker (2007, cf. also von Békésy 1947, Zwicker and Feldtkeller 1955, Hesse 1986). In the present case, the subjects listen alternately to the reference scene and the transmission system under consideration. The subjects’ task is to continuously adjust the input level of the transmission system so that it elicits the reference scene loudness. For that purpose, pairs of narrow-band impulses are used as test signals, one impulse presented in the reference scene, the other by the system under test. Typically, the reference scene is presented first, while the presentation order did not change the results of the evaluation experiment described below. After each pair, the center frequency is changed automatically. The reference scene level remains constant while the input level of the system under test is either increased or decreased with each frequency step. The subjects are asked to change the direction of the level variation using a hand switch every time the loudness of the two sounds in a pair differs. This procedure results in a frequency dependent zigzag-pattern, alternating around the level at equal loudness.

Two signal impulses with 0.4 s duration, 5 ms Gaussian gating according to Zwicker and Feldtkeller (1967, pp. 20–21), and 0.1 s spacing have proven useful as comparison pairs. To indicate the comparison pairs by the temporal stimulus organization, two successive pairs are separated by 0.4 s silence. A level variation with 1.5 dB step size, starting 10 dB above the level eliciting approximately the reference scene loudness, turned out advantageous. Each experimental run is divided in two consecutive parts, one with increasing center frequency, starting from 10 Bark (about 1.3 kHz, cf. section 2.3) upwards to 24.8 Bark (20 kHz), and one with decreasing frequency, starting from 12 Bark (1.7 kHz) downwards to 0.2 Bark (20 Hz), both with 0.05 Bark step size. On the CBR scale equidistant frequencies were selected to achieve an equally spaced auditory-adapted frequency distribution. In order to reduce methodical artifacts due to the beginning of the adjustment procedure, 20 steps of the results in the overlapping region between 10 and 11 Bark for the increasing and 12 and 11 Bark for the decreasing branch are dropped. The individual results are computed by interpolating the average levels of every two adjacent turning points.

Stimulus Selection and Verification of the Procedure Stimuli suitable for measuring LTFs are narrow-band signals, as for example pure tones or narrow-band noises (NBNs). To address the properties of both stimulus categories with regard to a reverberant reference scene[3], LTFs as defined above have been measured for a virtual acoustics system[4] with pure tones and half critical-band wide noises. The experiment was conducted with four experienced normal hearing[5] subjects utilizing the procedure introduced above, resulting in a duration of about 6 minutes for the noise and 11.5 minutes for the tone stimuli. Figure 2.8 shows the inter-individual median of the noise data (black contour) and the inter-individual 25% and 75% percentiles of the pure tone results (gray contours).

The results for both stimuli are qualitatively comparable, while the NBN data proceed due to the noise's broader spectral shape smoother, revealing less detail. The smoothing may be desirable, for example to suppress the visibility of spectrally narrow resonance effects. Further, the NBN stimuli allow reducing the measurement time by about 50%. Since a detailed system analysis showing spectrally narrow effects was attempted, LTFs were measured using pure tone stimuli in the course of this thesis. The procedure resulted for three experienced subjects repeating the experiment described in this paragraph in an inter- and intra-individually averaged deviation of ±2 dB between two runs on different days. The maximum intra-individual deviations between two runs did not exceed ±4 dB.

[3] Klein + Hummel Studio Monitor Loudspeaker O 98, setup according to section 5.4.2 in laboratory 1

[4] nonindividual dynamic binaural synthesis with Stax λ pro NEW headphones, average magnitude equalization, and Polhemus 3 Space FasTrack according to chapters 4 and 5

[5] pure tone threshold in quiet for Békésy-Tracking less than 20 dB above the long-term lab median

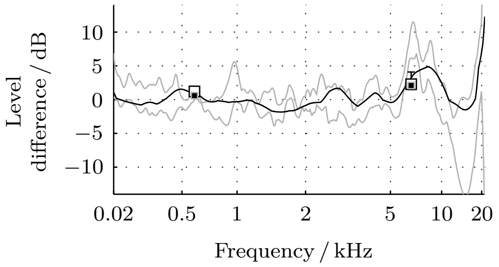

Figure 2.8: Loudness transfer function from a binaurally synthesized loudspeaker to the real counterpart. Median of individual tracking results measured with half critical-band wide noise (black contour) and 25% and 75% percentiles for pure tone test signals (gray contours). Further shown are quartiles of pure tone adjustment results at 650 Hz and 6 kHz with loudspeaker (open squares) and binaural synthesis reference (filled squares).

Figure 2.8 shows additionally for verification purposes the quartiles of the results of an adjustment experiment with pure tones at 600 Hz and 6.5 kHz in the same scenario. In that experiment, nine subjects adjusted the level of the signals presented by the system under test to the reference scene (open squares) and vice versa (filled squares). All subjects repeated each adjustment four times in random order, twice starting at a level notably above the target level, twice at a level clearly below. The adjustments were conducted in two sessions of 8 minutes average duration, one per frequency, with an intermediate break. The results confirm the applicability of the tracking procedure.

2.6 Just Noticeable Sound Changes

According to Fastl and Zwicker (2007, p. 175), the just audible physical change of a sound in general is referred to as just noticeable sound change (JNSC). In the course of this work, JNSCs are used to address the quality of electroacoustic transmission systems. JNSCs measured with an ideal transmission system must equal the JNSCs of the reference scene. This requirement is not sufficient for an ideal transmission system, but necessary. If spatial reference scene aspects are to be transmitted, the JNSCs related to spatial hearing are of interest. In this section, the procedures applied for JNSC measurements in this work are discussed along with the stimulus selection by the example of directional JNSCs.

Minimum Audible Angle

Mills (1958) defined the minimum audible angle (MAA) *"as the smallest detectable difference between the azimuths of two identical sources of sound"*, the angle formed

> *"at the center of the head by lines projecting to two sources of sound whose positions are just noticeably different, when they are sounded in succession."*

Mills' definition is extended to cover not only the azimuth, but also elevation and other source orientations, and to incorporate a distance requirement, resulting in definition 6.

Definition 6 (*Minimum Audible Angle*)

The angle between the lines from the center of the head to two stationary sound sources at the same distance with just noticeably different positions when sounded in succession is referred to as the minimum audible angle.

Earlier Studies Mills (1958) measured in an anechoic chamber the horizontal plane MAA for sources at 0.5 m distance with pairs of tone impulses of 1 s duration, 70 ms rise and fall time, and 1 s pause using a two-alternative forced choice (2-AFC) procedure. Every first impulse was presented from a reference position, while the second pulse was played back at different horizontal angular separations from the reference. The subjects had to indicate whether the second pulse originated left or right from the first. The MAA was taken as half the angular difference between the 25% and 75% points of the psychometric function obtained by a line fit to the proportion of responses "right" over the linearly scaled angle. The sources were positioned at angles symmetric around different reference sound incidence directions. According to Hartmann and Rakerd (1989), this procedure is equivalent to directly taking the angle where 75% of the responses are correct. Mills' results indicate dependencies of the MAA on the spectral stimulus content and on the sound incidence direction, with minimum values for frontal stimuli. At all sound incidence directions, the lowest MAAs arise for stimuli in the frequency ranges between 0.25 and 1 kHz as well as 3 and 6 kHz, large MAAs in the ranges around 1.5 and 8 kHz. The global minimum lies in the range of about 1°. Hartmann and Rakerd (1989) showed the procedure of Mills (1958) to be more likely an absolute than a relative localization task. They proposed a method referred to as two sources two intervals (2S2I) for addressing location discrimination, which compares two stimuli located symmetrically around the reference direction, without a source at the reference direction. However, Hartmann and Rakerd assume the MAA dependencies of frequency and direction found by Mills to be correct, since his procedure allows for comparing situations. This assumption is confirmed using a 2S2I method with critical-band wide NBN presented by wave field synthesis according to section 3.1.2 in a reverberant listening environment by Völk and Fastl (2011b).

Perrott and Pacheco (1989) studied the dependence of the MAA on the inter-stimulus interval for broadband pink noise impulses of 10 ms duration with temporal slopes shorter than 0.1 ms. The used adaptive 2-AFC procedure addressing whether the second hearing sensation occurred left or right from the first can be regarded as a variation of the 2S2I method of Hartmann and Rakerd (1989). The MAA was defined as the angular separation of two sources resolved at a 70.7% correct response level. The results indicate a decay of the frontal horizontal plane MAA from some 2.5° to about 1° when increasing the temporal stimulus separation from 1 ms to 150 ms, and a constant MAA for larger inter-stimulus intervals. Using a comparable procedure, Perrott and Saberi (1990) found, with pairs of 400 Hz click trains of 50 ms duration and 500 ms silence between the stimuli, frontal MAAs for sources in the horizontal plane of about 1° and for vertically spaced sources in the median plane of some 3.7°. For frontal sources in planes tilted from the horizontal, indicated by 0° tilting angle, in 10° steps to the median plane at 90° tilting angle, significant changes compared to 0° occur not before 80° tilting angle. These results were confirmed by Saberi et al. (1991) with broadband noise impulses of 250 ms duration and 500 ms pause. Perrott (1993) found a frontal horizontal plane MAA of about 1.2° for high-pass noise impulses with 1 kHz limiting frequency, 200 ms duration and pause with 10 ms slopes. For broadband noise impulses of 300 ms duration and pause with 10 ms slopes, Grantham et al. (2003) showed the frontal MAA to increase from about 1.6° in the horizontal plane to 2.8° in a diagonal plane, and to 6.5° in the median plane.

Procedure, Stimuli, and Verification Following Perrott and Pacheco (1989), an adaptive 2-AFC 2-down 1-up method according to Levitt (1971) is proposed for MAA measurements here. The stimuli to be compared are presented by sound sources at the same distance under head-related angles symmetric around the reference direction, and the step size is adapted by Parameter Estimation by Sequential Testing (PEST, Gelfand 2004). It is the subjects' task to indicate by pressing one of two buttons where the second hearing sensation occurred with regard to the first hearing sensation. This way, a criterion-free procedure is implemented (cf. Hellbrück and Ellermeier 2004, p. 224). The presentation sequence is chosen randomly and the procedure is repeated until both, the deviation between the last two minimum and the deviation between the last two maximum values are below a threshold value. The threshold is selected dependent on the stimulus, the playback system, and the source positions. Since the 2-down 1-up method converges to the 70.7% point of the psychometric function (Levitt 1971), the MAA is defined as the angular threshold where about 71% of all relative position judgments are correct. The adaptive procedure is repeated three times per stimulus and subject, and the intra-individual median per stimulus is taken as the individual result.

Based on the earlier studies, the minimum MAA is expected for frontal sound incidence and broadband impulses of more than 200 ms duration with an inter-stimulus interval exceeding 150 ms. According to Fastl and Zwicker (2007, p. 170), uniform exciting noise (UEN) provides constant intensity per critical band and is therefore assumed to be able to elicit all spectral localization cues. This being said, broadband UEN impulse pairs of 700 ms duration with 20 ms Gaussian gating according to Zwicker and Feldtkeller (1967, pp. 20–21) and 300 ms pause are selected to address system properties by MAAs.

In order to verify the adaptive procedure and the proposed method in general, the MAA for frontally incident plane waves simulated by wave field synthesis according to section 3.1.2 was measured for eight normal hearing subjects using the adaptive method described here and with a static 2-AFC method implementing the same task (cf. Völk and Fastl 2011b). In the static case, each angular separation was evaluated 20 times in random order per subject, ten times presenting the right source first, ten times the left. Figure 2.9 shows the inter-individual averages of the intra-individually averaged results.

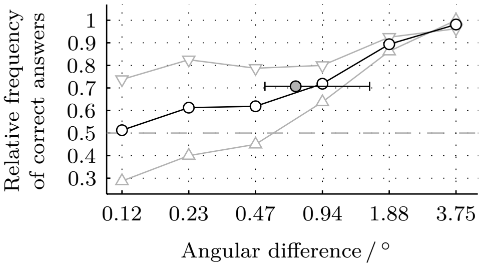

Figure 2.9: Average result of a static two-alternative forced choice minimum audible angle experiment (open circles) with broadband uniform exciting noise impulses presented as frontally incident plane waves approximated by wave field synthesis. Downward pointing triangles indicate the results for the stimulus sequence right source first, upward pointing triangles for the sequence left source first. Additionally depicted are the quartiles of the results from the corresponding adaptive procedure (filled circle with error-bars).

Displayed in figure 2.9 are statistics of all data (open circles) and separately the results for the presentation sequences right first (downward pointing triangles) and left first (upward pointing triangles). The median and inter-quartile range of the intra-individual medians representing the adaptive procedure are plotted horizontally (filled circle with error-bars). The results indicate a good agreement of static and adaptive procedure, since the adaptive procedure converges in good approximation to the 71% point of the average psychometric function. Comparing the results of the two presentation sequences indicates a tendency for the subjects to prefer the left button in the threshold region, where no or only little differences are audible. However, the transformed up-down method is designed robust towards response preferences (Levitt 1971), reflected in the good agreement of the average results of the static and adaptive procedures.

Averaging all horizontal MAA measurements conducted for frontal sound incidence in the course of this thesis results in approximately 0.7° for broadband UEN impulse pairs of 700 ms duration with 20 ms Gaussian gating and 300 ms pause (cf. Völk et al. 2010b, Schmidhuber et al. 2011, Völk and Fastl 2011b). This result agrees well with the earlier studies, further confirming the validity of the adaptive procedure.

The results of this section are summed up in that all studies discussed here report MAAs of about 1° for broadband stimuli presented in the frontal direction if the inter-stimulus interval exceeds 150 ms. For shorter inter-stimulus intervals, reduced bandwidth, and lateral sound incidence, the MAAs increase.

Minimum Audible Movement Angle

The MAA is measured with short sound impulses presented by two sources at fixed positions. If one source is moved during the measurement, the resulting angle increases dependent on the source velocity (Perrott and Musicant 1977, Grantham 1986). Following definition 7, this angle is referred to as the minimum audible movement angle (MAMA).

Definition 7 (*Minimum Audible Movement Angle*)

> *The angle between the lines from the center of the head to a moving and a stationary sound source at the same distance with just noticeably different positions when sounded in succession is referred to as the minimum audible movement angle.*

Perrott and Tucker (1988) as well as Chandler and Grantham (1992) report qualitatively comparable frequency dependencies of MAMAs and MAAs, with maximum values around 3 kHz. Chandler and Grantham found for an increase of the source velocity from 10°/ s to 180°/ s an increase of the frontal horizontal plane MAMA by the factor two, independent of the moving direction. According to Chandler and Grantham (1992), MAMAs and MAAs decrease monotonically with increasing stimulus bandwidth for all sound incidence directions and stimulus velocities. Changing the stimulus from a 3 kHz tone to broadband noise, the MAMA is reduced by approximately the factor ten. Regarding the source positions, Grantham et al. (2003) measured MAMAs in the horizontal, the median, and in diagonal planes, confirming the dependencies found earlier for MAAs and showing the MAMA to grow with increasing elevation.

Minimum Audible Distance

If two sources to be compared are positioned at different distances under the same head-related angle the JNSC with regard to distance perception is addressed. The result is according to definition 8 referred to as the minimum audible distance (MAD).

Definition 8 (*Minimum Audible Distance*)

> *The distance between two stationary sound sources on a line through the center of the head with just noticeably different positions when sounded in succession is referred to as the minimum audible distance.*

Edwards (1955) reports the accuracy of distance judgments to decrease with the source distance, also supported by data of Völk (2010b). Laws (1972a,b) found the MAD to increase with the source distance, later confirmed by Völk et al. (2012c), who report average MADs between 0.05 m and 1 m, for source distances between 0.5 m and 10 m.

The method proposed here for MAD measurements is adapted from the MAA measurement procedure described above. Differences are that the stimuli to be compared are presented by sound sources positioned symmetrically around a reference *distance* and the subjects' task: The participants are asked to indicate by pressing one of two buttons if the second hearing sensation occurred farther or closer than the first hearing sensation. The stimuli proposed for the system evaluation by MADs are the UEN impulses also used for the system evaluation by MAAs.

Figure 2.10 shows the inter-individual averages of the intra-individually averaged results of eight normal hearing subjects for an exemplary MAD measurement with broadband UEN impulse pairs of 700 ms duration with 20 ms Gaussian gating and 300 ms pause. The stimuli were presented by frontal wave field synthesis point sources according to section 3.1.2 at distances to the center of the head distributed symmetrically around 10 m.

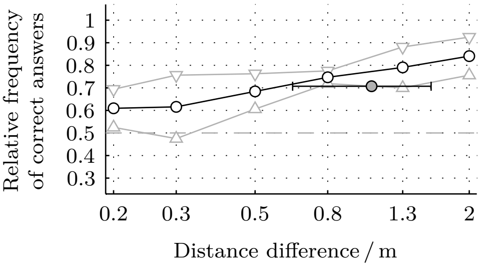

Figure 2.10: Average results of a static two-alternative forced choice minimum audible distance experiment (open circles) with broadband uniform exciting noise impulses played back by frontal wave field synthesis point sources at distances around 10 m. Downward pointing triangles indicate the stimulus sequence closer first, upward pointing triangles indicate the sequence farther first. The filled circle indicates the result of the corresponding adaptive procedure.

The experiment was conducted using the adaptive procedure proposed here and with a static two-alternative forced choice method implementing the same task in a reverberant laboratory (6.8 m × 3.9 m × 3.3 m, 50 ms average reverberation time, cf. Völk et al. 2012c). In the static case (open symbols), each source configuration was evaluated by all subjects 40 times in random order (open circles), 20 times presenting the closer sound first (downward pointing triangles) and 20 times the farther sound first (upward pointing triangles). The filled circle with horizontal error-bars indicates median and inter-quartile range of the intra-individual median of three repetitions of the adaptive procedure per subject.

The results shown by figure 2.10 indicate that the adaptive procedure employed tends to overestimate the MAD defined by the 70.7% point of the psychometric function. The results of the static procedure indicate a dependence of the MAD on the presentation sequence. It may be concluded that a change of the source distance away from the listener, indicated by the downward pointing triangles, is more likely detected. This effect is possibly accompanied by a tendency of the listeners to prefer the farther button, also indicated by the rather shallow psychometric function within the distance differences covered, not reaching 100%. However, both effects cannot be separated based on the data. The adaptive procedure converges on average to the static situation with the farther sound presented first that is the larger MAD. The fact that the transformed up-down procedure is designed robust towards response preferences (Levitt 1971) may provide further support to the assumption of two MADs at a given distance, depending on the presentation sequence. Evaluating the performance of an audio transmission system is possible using the adaptive method introduced, if the procedure converges to the same point of the psychometric function in the reference scene and the playback situation.

2.7 Summary

This chapter is initiated by the clarification of terms and definition of variables. Further, refined analytic formulae for the frequency dependence of critical bandwidth and critical-band rate are proposed. The critical bandwidth function converges to zero at 0 Hz, in contrast to current formulae, while fitting the original listening experiment results better than the established formulae. These properties allow for implementing signal processing routines based on the critical-band concept with reduced low-frequency artifacts. The critical-band rate function is in contrast to the established formula invertible and approximates the original listening experiment results more accurately, especially in the frequency range above 16 kHz. The refined definitions are employed to derive a procedure referred to as auditory-adapted analysis (AAA), primarily intended for *system* analysis, but covering *signal* analysis as well. With regard to the visualization of the results, the AAA spectrogram is introduced, combining auditory relevant magnitude and phase information in a single spectrogram. Further, a classification of acoustic signal processing systems as spectrally or spectro-temporally effective systems is proposed. Spectrally effective systems are fully represented by auditory-adapted magnitude transfer functions and group delays, in combination referred to as auditory-adapted transfer characteristics. Spectro-temporally effective systems are represented by AAA spectrograms. For perceptual system analysis, loudness transfer functions (LTFs) are defined, and directionally just noticeable sound changes revised with regard to the quality evaluation of audio signal processing systems. The procedures, tools, and definitions introduced in this chapter provide the methodical and conceptual basis for a psychoacoustically motivated discussion of virtual acoustics systems, as shown by the example of binaural synthesis in chapters 4 and 5.

3 Conjunctions of Virtual Acoustics and Hearing Research

Initially in this chapter, a *classification scheme for approaches to virtual acoustics* based on the respectively underlying concept is proposed. Using this classification scheme, an overview of current methods for the generation of virtual acoustical environments is given, motivating the selection of binaural synthesis as the exemplary method discussed in detail in the subsequent chapters. Furthermore, a reflection on psychoacoustic methods in general and the *stimulus definition* in detail serves as basis for the discussion of interrelations between virtual acoustics and hearing research. Especially the *application of psychoacoustic methods to the quality evaluation of virtual acoustics systems* and the *employment of virtual acoustics as playback method for hearing research* illustrate the necessity of a thorough theoretical framework as the basis of virtual acoustics. In that context, *issues in conventional headphone reproduction and related equalization procedures* are identified and discussed with respect to virtual acoustics, especially including a revision of the application range of free-field equalization. Aspects regarding the quality evaluation of virtual acoustics systems conclude the chapter.

3.1 Categorization of Approaches to Virtual Acoustics

According to definition 1, eliciting specific hearing sensations is the final goal of virtual acoustics (VA). Two fundamentally different approaches to the generation of predefined sensations are discussed:

1. Application of stimuli known to trigger under defined conditions approximately the hearing sensations intended, making use of knowledge about the connections between stimuli and sensations (*psychoacoustic* relationships, cf. section 3.2.1 and Fastl and Zwicker 2007, p. 11). The degree of authenticity achievable with these so-called *psychoacoustically motivated approaches* depends on the situation to be simulated, the psychoacoustic relationships employed, the actual implementation, the listening environment and conditions, and the listener.

2. Recreation of the acoustical reference scene stimuli (the *physical* signals acoustically contributing to eliciting the hearing sensations to be generated, cf. section 3.2.2 and Fastl and Zwicker 2007, p. 11). If successful, these physically motivated approaches theoretically trigger under the non-acoustically not modified reference scene conditions for the same subject the sensations intended. However, inter-modal, cognitive, and memory effects may influence hearing sensations (cf. section 3.2).

Based on the fundamental principle, VA systems are assigned to the classes of physically and psychoacoustically motivated approaches in the following. Initially, the classes are discussed in detail, along with the definitions required. For affecting all playback situations, inter-modal, cognitive, and memory effects are neglected in the classification. However, the amount of non-acoustic influences on hearing sensations may depend on the VA system used (Völk et al. 2010b).

3.1.1 Psychoacoustically Motivated Approaches to Virtual Acoustics

The class of psychoacoustically motivated approaches to VA is identified in that the acoustic stimuli of the reference scene are not of direct interest in the reproduction system design. The reproduction system is rather developed based on previously acquired psychoacoustical knowledge about the perceptions expected to be elicited by the stimuli employed *in the reproduction situation.* Consequently, recreating the original sound field is not the goal of psychoacoustically motivated approaches to VA (definition 9).

Definition 9 (*Psychoacoustically Motivated Approaches*)

Psychoacoustically motivated approaches to virtual acoustics are based on knowledge about the connections between stimuli and hearing sensations to select stimuli eliciting the hearing sensations to be generated as accurately as possible.

Typically, psychoacoustically motivated VA systems work best within a certain spatial region (e. g. Warncke 1941, Aoki et al. 1990). This region is usually referred to as the spatial *sweet spot* (definition 10, Rose et al. 2002, Dickreiter 2003, Eargle 2005).

Definition 10 (*Sweet Spot*)

A specific audio reproduction system works best for listeners situated within a system dependent spatial region referred to as the system's (spatial) sweet spot.

By definition, recording for psychoacoustically motivated approaches cannot follow physically thorough, mathematically exactly defined rules, but in fact resembles an artistic process, requiring the knowledge and creativity of the recording and production engineers (e. g. Snow 1955, Olson 1977, Kahana et al. 1999, Dickreiter 2003). This circumstance often results in a shift of the design target from reproducing the original hearing sensations authentically to creating a convincing, desired, or impressive playback scenario (cf. Klipsch 1960, Bernfeld 1975, Furuya et al. 2004, Eargle 2005, Webers 2007). A fact to be considered for the definition of design targets for audio reproduction systems in general is the possible difference between the authentic situation and the most plausible or most convincing playback scenario (Theile 1991, 2001, Larsson et al. 2008). Furthermore, psychoacoustically motivated approaches to VA usually involve a post-production step (Dickreiter 2003, Eargle 2005, Webers 2007), further modifying the recorded signals with the aim of eliciting more convincing or pleasant, not necessarily more authentic hearing sensations (e. g. Aoki et al. 1990, Bech 1998, Zarouchas and Mourjopoulos 2009).

Stereophonic Audio Reproduction The currently best established psychoacoustically motivated approaches to VA belong to the class of *stereophonic audio-reproduction systems* (Warncke 1941, Fletcher 1941, Snow 1954, 1955, Theile 1981, Dickreiter 2003). The first stereophonic sound transmission took place in 1881 between the Théâtre Français, the Paris Opera, and the Palais de l'Exposition, the exhibition hall of the first Electric Exhibition in Paris, France (for a historic overview cf. Hertz 1981). Based on extensive research by the group around Harvey Fletcher at Bell Labs, New York, USA, stereophonic recordings became available in the 1920s. In 1933, the first long range stereophonic live transmission took place between Philadelphia and Washington, D. C., documented in a series of papers in the Bell System Technical Journal (Fletcher 1934, Steinberg and Snow 1934, Wente and Thuras 1934, Scriven 1934, Affel et al. 1934, Bedell and Kerney 1934).

The most simple standardized stereophonic reproduction setup consists of two loudspeakers (LSs) positioned at the corners of an equilateral triangle with the listener at the third corner (Snow 1955, Roys 1977, ITU-R BS.775-2 2006). This setup can be found in slightly modified incarnations in traditional living room HiFi installations (Theile and Plenge 1977, Theile 1978, 1980), television sets (Sippl 1988), or budget class car audio systems (Ashley 1976, Azzali et al. 2004). Stereophonic reproduction is based on the principle of *summing localization*, occurring for example if two sound sources radiate coherent signals in specific geometric arrangements of sources and listener, for example at the corners of an equilateral triangle. For level differences and time offsets between the source signals, typically one hearing sensation referred to as phantom (sound) source arises at a position between the sources (definition 11, Warncke 1941, Snow 1954, Platte and Genuit 1980, Zollner and Zwicker 1993, Blauert 1997, p. 204).

Definition 11 (*Stereophonic Audio Reproduction*)

> *Two-channel audio reproduction systems based on the summing localization of two loudspeakers are referred to as stereophonic audio reproduction systems.*

It is possible to vary the phantom source position within certain limits by the level difference and temporal offset between the source signals (in horizontal direction, Leakey 1959, Klipsch 1960, Blauert 1997, p. 206), the overall level (distance, Zahorik et al. 2005), and spectral weighting, making use of a psychoacoustical phenomenon called directional bands (distance and vertical direction, Blauert 1997, pp. 108–116). In general, summing localization is more likely to occur for frontal rather than lateral sound sources (Theile and Plenge 1977), and summing localization is more stable for sources in the horizontal plane rather than for other sources (Pulkki 2001, Barbour 2003). The position of the phantom source depends on the geometric situation (Blauert 1997, p. 218), the listening room conditions (Aoki et al. 1990, Benjamin 2006, Kim et al. 2008), and the LS radiation characteristics (Bauer 1960, Keele 1983, Davis 1987, Aarts 1993, Toole 2000). Current efforts aim at continuous adaptation of the stereophonic sweet spot for fixed LS positions to variable listener positions (Merchel and Groth 2009a,b, Groth and Merchel 2009) and head rotations (Merchel and Groth 2010a,b) by preprocessing the LS input signals based on head tracking data.

Since stereophonic reproduction systems have been in use for more than a century, recording and production engineers, technicians, and managers are accustomed to the associated well-established production processes and the resulting auditory impressions, typically designed to sound most convincing or most impressive (Theile 1991). That being said, the auditory impressions elicited by an authentic reproduction system and the associated production processes are likely considered unusual or, compared to the established approaches, unnatural. This human factor must be taken into account when expert judgments, especially on physically based approaches to VA (cf. section 3.1.2), are discussed or when new approaches are to be introduced or standardized.

Surround Sound Reproduction Audio playback systems usually referred to as *surround sound reproduction systems* (ITU-R BS.775-2 2006, Meares and Theile 1998, Couling 1999) belong to the class of psychoacoustically motivated approaches to VA. Such systems are typically based on stereophonic principles, employ five, seven, or more channels with independent signals and at least as much LSs, and are in use commonly in home theater setups or venues like cinemas or theaters (definition 12, Silzle and Theile 1990, Theile 1990, O'Dwyer et al. 2004, Faller 2006). Current activities lean towards including additional LSs above the horizontal plane with the aim of eliciting hearing sensations at elevated positions (Furuya et al. 2004, Hamasaki et al. 2005, Sundaram and Kyriakakis 2005, Ehret et al. 2007, Lee et al. 2010, Lee 2011).

Definition 12 (*Surround Sound Reproduction*)

> *Multi-channel audio reproduction systems with more than two channels based on summing localization are referred to as surround sound reproduction systems.*

Surround sound systems can be regarded as a combination of multiple stereophonic systems at different directions with regard to the listener. For that reason, the principles of stereophonic systems also apply to surround sound systems (Zieglmeier and Theile 1996), while the psychoacoustical laws of summing localization are different for more than two sources (Blauert 1997, p. 275). In general, stable phantom sources are possible in front of the listener, whereas the lateral and elevated application of stereophonic techniques increases the diffuseness of the hearing sensations while not resulting in stable phantom sources (Theile and Wittek 2011).

Vector Base Amplitude Panning (VBAP, Pulkki 1997, Pulkki and Karjalainen 2001, Pulkki 2001) also belongs to the class of surround sound systems. VBAP playback adjusts the source signal levels for arbitrary source configurations using vector based rules (Pulkki 1997). While the approach at first glance seems to extend the possibilities of stereophonic reproduction, it is necessarily limited to the psychoacoustical dependencies of summing localization for time aligned signals (Pulkki and Karjalainen 2001, Pulkki 2001, Batke and Keiler 2010). Mathematic formulae for the prediction of hearing sensation positions also exist for stereophonic systems (e. g. Bernfeld 1975).

Conventional Headphone Reproduction Conventional headphone (HP) reproduction may also be regarded as a psychoacoustically motivated approach to VA. The resulting hearing sensations are usually located within the head (in the head localization, Fastl and Zwicker 2007, p. 308). If the same signal is fed to both HPs, typically a hearing sensation occurs in the middle of the head (Fastl and Zwicker 2007, figure 15.12). By introducing a level difference $\Delta L < 30\,\mathrm{dB}$ or a temporal offset $\Delta\tau < 0.6\,\mathrm{ms}$ between the signals fed to the HPs, the hearing sensation position can be shifted laterally within the head towards the HP driven by the higher level or earlier signal (lateralization, cf. Jeffress and Taylor 1961, Toole and Sayers 1965, Wallerus 1976). Even larger level differences result in a single hearing sensation position close to the ear receiving the higher level; for larger temporal offsets, two hearing sensation positions occur (Blauert 1997, p. 224). It is possible to combine level difference and temporal offset, if correctly parameterized resulting in more stable hearing sensation positions (Plenge 1974). In HP playback, it is possible to employ only one capsule for reproduction (monotic playback), to drive both capsules with the same signal (diotic playback), or to drive each capsule with a different signal (dichotic playback, nomenclature according to Stumpf 1905, cf. definition 13).

Definition 13 (*Monotic, Diotic, and Dichotic Headphone Playback*)

> *Monotic headphone playback denotes stimulation of one ear only. When using two headphone capsules to stimulate both ears, it is possible to feed both capsules with the same signal (diotic playback) or with different signals (dichotic playback).*

The similarity between HP playback and stereophonic reproduction as described above leads to lateralized instead of localized spatial distributions of hearing sensations when playing back stereophonic signals by HPs. For most audio content, hearing sensation positions within the head are not authentic. Based on an informal survey conducted in the course of this thesis, this fact is widely tolerated and often goes unnoticed, especially by audio consumers who never thought consciously about hearing sensation positions. The popularity of listening with HP presently grows based on the increasing use of mobile communication and audio playback devices.

By definition, also the playback methods free-field and diffuse-field equalized headphone reproduction belong to the class of psychoacoustically motivated approaches to VA (cf. Pritchett 1954, Zwicker and Maiwald 1963). The procedure later referred to as free-field equalization was proposed by Fletcher and Munson (1933) and described explicitly by Beranek (1949, p. 730, equal loudness method) as well as Zwicker and Gässler (1952). Narrow-band stimuli presented by free-field equalized HPs elicit the same loudness as a free sound field (Fastl and Zwicker 2007, p. 8). According to Bocker and Mrass (1959), the preferable procedure for acquiring the target curve for the free-field equalization of specific HPs is to acquire their inter-individually averaged loudness transfer function with respect to a frontally incident narrow-band plane wave reference of known sound pressure level (cf. definition 5). Since it is important to listen binaurally to the plane wave and diotically to the HPs, the subjects have to take off the HPs to listen to the reference field (Bocker and Mrass 1959). Diffuse-field equalization denotes the analogous procedure with a diffuse-field reference (Theile 1981). Diffuse-field equalized HPs are often considered more euphonious and better suited for the playback of stereophonic recordings than free-field equalized models (Theile 1986, 1991). This preference is presumably due to the boost of the so-called presence band in the range of 3 kHz in the diffuse-field compared to the free field (Fastl and Zwicker 2007, figure 8.2). A presence-band accentuation is known to increase speech intelligibility and voice quality (Allen et al. 1969, Blauert 1970).

Stimuli can be presented by equalized HPs at a free-field or diffuse-field equivalent level equal to the reference field level. The equivalent level *must not* be regarded as the actual sound intensity or sound pressure level in the HP listening situation, as occasionally done (e. g. Schorer 1986, Fastl and Zwicker 2007, p. 308). Reference scene and equalized HP presentation are related only in that the reference sounds presented by free-field or diffuse-field equalized and calibrated HPs elicit, for a typical subject, the reference scene loudness (cf. section 2.5). Due to this perceptual relation of playback situation and reference scene, free-field and diffuse-field equalization belong to the class of psychoacoustically motivated approaches to VA. The classification of conventional HP playback can be generalized to conventional playback with two sound sources close to the ears, including the playback of stereophonic signals by per se physically motivated approaches as for example crosstalk cancellation, which is discussed amongst others in the following section.

Other Psychoacoustically Motivated Approaches Most other psychoacoustically motivated approaches to VA can be attributed to one or a combination of the procedures discussed above (e. g. König 1994, 1995, Shimada and Hayashi 1995, Kahana et al. 1999, Baumgarte and Faller 2003, Martignon et al. 2005, Kim et al. 2008). For that reason, no other approaches are described in detail in this section.

3.1.2 Physically Motivated Approaches to Virtual Acoustics

It is the aim of physically motivated approaches to VA to create exactly the stimuli of the reference scene. In other words, the generated sound field is meant to match the reference field at least in a specified spatial region (definition 14). Whether this goal is reached completely or partially is irrelevant for the classification of an approach.

Definition 14 (*Physically Motivated Approaches*)

Physically motivated approaches to virtual acoustics aim at recreating the physical sound stimuli eliciting hearing sensations in the reference scene.

Typically, correct reproduction is intended in a limited spatial region referred to as listening volume. If the target region for correct reproduction is limited to a plane, it is denominated as listening area (definition 15).

Definition 15 (*Listening Area and Listening Volume*)

A physically motivated approach to virtual acoustics attempts synthesis in a spatial region, which is referred to as listening area if the reproduction is targeted in a plane or as listening volume if the reproduction is optimized for a volume.

Binaural synthesis (BS), a physically motivated approach to virtual acoustics aiming at correctly reproducing the sound pressures at the eardrums, is a major concern of this thesis and is therefore revised theoretically in chapters 4 and 5. An overview of other physically motivated approaches and the motivation for selecting BS as the exemplary procedure studied here are given in the following.

Wave Field Synthesis *Wave field synthesis (WFS)* was developed in the 1980's by the group around A. J. Berkhout at Delft University of Technology (Berkhout 1988, Berkhout and de Vries 1989). WFS is an audio playback procedure aiming at synthesizing the reference sound field in a listening area (Berkhout et al. 1993). Theoretically, the synthesis is possible in a three-dimensional listening volume based on the Kirchhoff-Helmholtz integral equation, using an infinite number of secondary monopole and dipole point sources distributed continuously over the listening volume boundary area (Vogel 1993).

Definition 16 (*Virtual, Primary, and Secondary Sound Sources and Fields*)

> *The virtual sources to be simulated by means of virtual acoustics are referred to as virtual or primary (sound) sources, generating respective virtual or primary (sound) fields. Secondary (sound) sources generating secondary (sound) fields are used by a virtual acoustics system for playback in the reproduction situation.*

For practical implementations, LSs are employed as secondary sources (cf. definition 16) and the correct reproduction is usually attempted in a two-dimensional listening area (Verheijen 1997). Therefore, the primary fields are limited to two-dimensional fields independent of the coordinate direction orthogonal to the listening area, and errors occur in the synthesized field (Völk et al. 2011b). The errors result from the reduction to a listening area (Spors et al. 2008) and the deviation of the LS radiation characteristics from point source, monopole, and dipole characteristics (Zollner and Zwicker 1993, Start 1997). In addition, the synthesis procedure is typically reduced to monopole secondary sources, resulting in an erroneous wave field outside the listening area that possibly propagates directly or reflects back into the listening area, where it distorts the intended sound field. It is not possible to continuously distribute infinitely small LSs on a surface. For that reason, spatial sampling is necessary, resulting in artifacts since the sampling theorem is not fulfilled in practical situations (Wittek and Augustin 2005, Wittek 2007, Wittek et al. 2007a,b). Revisions of the WFS theory are given by Völk et al. (2011b) and Völk and Fastl (2012), perceptual aspects are addressed for example by Völk (2010b), Lindner et al. (2011), Schmidhuber et al. (2011), Völk et al. (2010c), and Völk et al. (2012c). The simulation of WFS by binaural synthesis, referred to as *Virtual Wave Field Synthesis (VWFS)*, is discussed by Völk et al. (2008b, 2010a).

Ambisonics In 1973, Gerzon introduced a recording and playback method for audio reproduction in three dimensions termed *Periphony* (now *Ambisonics*), employing four or more LSs distributed at different angles in azimuth and elevation around a listener. Each LS driving signal is recorded in the reference scene with a directional microphone facing the respective LS position. Perfect reproduction would theoretically be possible with ideally unidirectional microphones and an infinite number of idealized secondary sources distributed on a sphere around the listener (Gerzon 1973). Systems with a finite number of channels provide reduced directional resolution. The Ambisonics principle is elegantly described using spherical harmonics of different orders, mathematically proving the obtainable directional resolution to be proportional to the number of LSs (Leitner et al. 2000, Ahrens and Spors 2008a).

In its most simple implementation, Ambisonics is designed as a first-order system with three or four LSs (Nicol and Emerit 1998), while current research leans towards more channels, referred to as *higher-order Ambisonics* (HOA, Ahrens and Spors 2008b, Trevino et al. 2010, Clapp et al. 2010). Ambisonics has been extended beyond recording and playback to the synthesis of sound fields including directional (Ahrens and Spors 2007) and focused sound sources (sources within the LS setup, Ahrens and Spors 2008c), as well as plane waves (Ahrens and Spors 2008d). In the idealized case of a continuous spatial source distribution, Ambisonics can be regarded as a specific WFS configuration, while the approaches show, due to the different theoretical derivations, diverging advantages and disadvantages in real implementations (Nicol and Emerit 1998, Daniel et al. 2003). Detailed comparisons of Ambisonics and WFS reveal the spatial structure of the artifacts resulting from the discretization of the secondary source distribution as a major difference (Spors and Ahrens 2007), with WFS showing advantages regarding the audibility of the spatial aliasing for the synthesis of plane waves (Spors and Ahrens 2008). Sound field simulations indicate HOA to be more robust regarding the area of correct reproduction and computationally more efficient for sound field synthesis than WFS, while the approaches show similar limitations regarding the recording processes (Daniel et al. 2003).

Simulated Open Field Environment Seeber et al. (2010) proposed a multi-LS audio synthesis procedure referred to as *Simulated Open Field Environment (SOFE)*, capable of approximating the spatio-temporal reflection pattern arising at a specific position of the reference scene at the optimization position in an anechoic chamber based on the Ambisonics principle. Single reflections are computed by a finite order mirror-source model and played back by the respectively nearest LS or a combination of two neighboring LSs, using a stereophonic approach (cf. section 3.1.1) discussed in detail by Seeber and Hafter (2007). This procedure allows for controlling the reflection pattern by means of single reflections, which can be helpful in hearing research. A prerequisite for the SOFE to produce valid results is the system operation under anechoic conditions so that no reproduction room reflections occur. Since the SOFE aims at reconstructing the field at an optimization position, it belongs to the class of physically motivated approaches to VA.

Crosstalk Cancellation *Crosstalk cancellation (CTC)* denotes a procedure aiming at generating predefined sound pressure signals at two or more spatial optimization points. Therefore, it is intended to drive two or more sound sources in such a way that the signal at each optimization point can be controlled independently from the signals at all other optimization points (Atal et al. 1966, Cooper and Bauck 1989). Static CTC systems are designed for fixed source and optimization positions, whereas dynamic systems adapt themselves at the runtime to variable geometrical configurations. Exemplary, the generation of two independent sound pressure signals by static CTC at artificial head microphones located at the eardrum positions is regarded, since this procedure is frequently applied to real heads for presenting binaurally synthesized signals without HPs (cf. Møller 1989, Köring and Schmitz 1993). In this configuration, CTC combined with BS represents a physically motivated approach to VA. While HP based BS is the main focus of this thesis, CTC is discussed since it allows, combined with the BS framework derived in chapters 4 and 5, for the system-theoretically correct LS based ear signal reproduction for static listeners in anechoic environments. If the sound pressure signals are represented by the microphone output voltages u_L and u_R, the aim of positioning the signals represented by $u_\mathrm{e,L}$ and $u_\mathrm{e,R}$ at the eardrums is described in the frequency domain by

$$U_\mathrm{e,L} \overset{!}{=} U_\mathrm{L} \qquad \text{and} \qquad U_\mathrm{e,R} \overset{!}{=} U_\mathrm{R}. \tag{3.1}$$

The most simple static CTC approach employs two LSs with the voltage spectra $U_\mathrm{ls,L}$ and $U_\mathrm{ls,R}$ at their input terminals. In this case, the transfer functions (TFs)

$$\begin{aligned} H_\mathrm{LL} &= U_\mathrm{e,L}^\mathrm{ls,L}/U_\mathrm{ls,L}, & H_\mathrm{RL} &= U_\mathrm{e,L}^\mathrm{ls,R}/U_\mathrm{ls,R}, \\ H_\mathrm{RR} &= U_\mathrm{e,R}^\mathrm{ls,R}/U_\mathrm{ls,R}, \quad \text{and} & H_\mathrm{LR} &= U_\mathrm{e,R}^\mathrm{ls,L}/U_\mathrm{ls,L} \end{aligned} \tag{3.2}$$

describe the playback situation, with $U_\mathrm{e,L}^\mathrm{ls,L}$ denoting the contribution of the left LS to the left ear signal spectrum and so on. Due to their fundamental role in the CTC process, the TFs given by equation 3.2 are commonly referred to as CTC filters. The spectra at the eardrums resulting from the LS input signals $u_\mathrm{ls,L}$ and $u_\mathrm{ls,R}$ can be written as

$$\begin{aligned} U_\mathrm{e,L} &= U_\mathrm{e,L}^\mathrm{ls,L} + U_\mathrm{e,L}^\mathrm{ls,R} = U_\mathrm{ls,L} H_\mathrm{LL} + U_\mathrm{ls,R} H_\mathrm{RL} \quad \text{and} \\ U_\mathrm{e,R} &= U_\mathrm{e,R}^\mathrm{ls,R} + U_\mathrm{e,R}^\mathrm{ls,L} = U_\mathrm{ls,R} H_\mathrm{RR} + U_\mathrm{ls,L} H_\mathrm{LR}. \end{aligned} \tag{3.3}$$

Consequently, the spectra of LS input signals fulfilling the requirements imposed by equation 3.1 are given by

$$U_\mathrm{ls,L} = \frac{U_\mathrm{L} H_\mathrm{RR} - U_\mathrm{R} H_\mathrm{RL}}{H_\mathrm{RR} H_\mathrm{LL} - H_\mathrm{RL} H_\mathrm{LR}} \quad \text{and} \quad U_\mathrm{ls,R} = \frac{U_\mathrm{R} H_\mathrm{LL} - U_\mathrm{L} H_\mathrm{LR}}{H_\mathrm{RR} H_\mathrm{LL} - H_\mathrm{RL} H_\mathrm{LR}}. \tag{3.4}$$

If the geometric arrangement of the recording situation, especially LS position, head position, and room properties, is constant, driving LSs with the signals $u_\mathrm{ls,L}$ and $u_\mathrm{ls,R}$ results in microphone output signals at an evaluation artificial head equal to u_L and u_R.

The generic CTC procedure can be extended to more than two CTC sources (Hokari et al. 2001, Huang et al. 2007) and more than two receivers (Bauck and Cooper 1996, Kahana et al. 1997). Since closed mathematical solutions are not possible in these cases, different approximations have been proposed (Bauck and Cooper 1996, Kirkeby and Nelson 1999, Norcross et al. 2004a). In general, due to the TF inversions in equation 3.4, high linear amplification is required, resulting in a reduced overall dynamic range and high sensitivity to changes in the room reflection patterns. Frequently used approaches of reducing implementation problems resulting from the TF inversions for a fixed source configuration are regularization (Kirkeby et al. 1998, Mouchtaris et al. 2000), joint least squares optimization (Ward 2000, Rao et al. 2007), Wiener filtering in the time domain (Kim and Wang 2003), or p-norm optimization (Jungmann et al. 2011). An approach modifying the geometrical source distribution reproducing higher frequencies in front of the subject and lower frequencies more laterally, referred to as Optimal Source Distribution (OSD), was proposed by Takeuchi and Nelson (2002). Since this procedure theoretically requires continuous source and frequency range distributions, in practical implementations discretization procedures are necessary (Bai et al. 2005, Takeuchi and Nelson 2007, Akeroyd et al. 2007). In section 5.1.5 of this work, an approach referred to as auditory-adapted exponential transfer function smoothing (AAS) is introduced, aiming at imperceptible spectral smoothing (cf. Völk et al. 2011a). AAS applied to the denominators of equation 3.4 is capable of reducing implementation problems resulting from the TF inversion in a perceptually motivated manner.

In general, small geometric displacements may under anechoic or slightly reverberant circumstances be tolerated (Takeuchi et al. 2001), but freely moving receivers deteriorate the CTC (Ward and Elko 1999). Dynamic systems adapt the CTC based on receiver position data detected by a tracking device. Gardner (1997) presented a dynamic CTC system with two LSs, optimized for positioning two sound pressure signals at the eardrums, and working as intended for viewing directions within the angle spanned by the head and the LSs. Outside this area, artifacts like sound coloration or ringing occur due to filter inversion problems (Gardner 1997). Lentz (2006) proposed an approach using four LSs mounted above ear height (cf. Lentz and Schmitz 2002, Lentz and Behler 2004), which adaptively selects the combination of two LSs with the best possible CTC at that moment, predicted according to Köring and Schmitz (1993). If in a specific situation multiple configurations allow for the best case result, a cross-fade between two neighboring CTC systems takes place. Menzel et al. (2005, 2006) introduced a system using focused WFS sources as CTC sources, which are located at fixed positions with regard to the listener's head. The major shortcomings of this approach are the high amount of hardware required and the spatial aliasing of the WFS, producing audible artifacts (Boone et al. 1995). A concept that could overcome the aliasing issues and hardware effort by employing phantom sources as the CTC sources is introduced and discussed by Völk et al. (2009a).

Aspects of the Scene Representation Regarding Storage and Transmission Considering storage and transmission of VA scenes, a difference between psychoacoustically and physically motivated approaches arises. Typically, virtual scenes are described by physical parameters as for example listening room characteristics, sound source radiation patterns, and geometrical scene and listener configurations. In the context of virtual environments or software development, this physical description is referred to as an object based scene representation (Isdale 1993). Physically motivated approaches to VA straightforwardly allow implementing object based scenes since they attempt to control the physical sound field and therefore provide the respective parameters. Consequently, physical objects are advantageous in the scene coding for physically motivated approaches to VA. This is reflected in that frequently applied strategies for coding and transmitting scenes for physically motivated approaches to VA are based on physical parameters (Horbach and Boone 1999, Faller 2003, Engdegård et al. 2008). Psychoacoustically motivated approaches to virtual acoustics on the contrary provide no direct relation of physical sound field parameters and the synthesis procedure. As a consequence, no direct way of implementing physical objects in psychoacoustically motivated approaches to virtual acoustics exists. The strict mathematical implementation of sound sources in psychoacoustically motivated approaches based on physical object descriptions may result in undesired side effects, as for example the coloration artifacts often associated with excessively boosting directional bands (Blauert 1997, pp. 108–116). For that reason, scenes to be reproduced by psychoacoustically motivated approaches to VA are frequently encoded channel based, with discrete channels assigned to the electroacoustic transducers of a specifically arranged playback setup. The scene is encoded in the transducer input signals, typically produced by a so-called mixing process, which means expert listeners adjust the signals by auditory control using the specified playback setup (Benjamin 1998, ITU-R BS.775-2 2006).

A rule based alternative to the discrete channel based encoding is the definition of perceptual objects, as for example hearing sensations, characterized by a set of hearing sensation properties. This way, an object based scene representation is possible also for psychoacoustically motivated approaches to VA, especially advantageous for storage and transmission. However, the actual physical rendering by psychoacoustically motivated approaches in the playback situation still suffers from the side effects discussed above. Perceptual objects used for storing and transmitting scenes for psychoacoustically motivated approaches to VA could be extended to cover scenes for physically motivated approaches. However, this procedure would require an additional possibly lossy preprocessing step, transcoding the virtual scene descriptors from physical to perceptual objects.

Motivation of Selecting Binaural Synthesis as the Exemplary Approach Interrelations between VA and hearing research are discussed using BS as the exemplary approach studied in detail. A physically motivated approach to VA is selected since this class of approaches is suited for the audio playback in hearing research (cf. section 3.2.3). Further, an approach not requiring a free-field environment to work as intended is desired. Since all LS based physically motivated approaches discussed in this section require in practical implementations free-field conditions to work correctly, the choice of BS, the only well-established HP based physically motivated approach, becomes obvious.

3.2 Procedural Aspects and the Audio Playback in Hearing Research

The fundamental concern of Psychoacoustics is establishing relations between physical stimulus properties and the corresponding hearing sensations (Fechner 1860, Stevens and Davis 1938, Feldtkeller and Zwicker 1956, for an overview cf. Fastl and Zwicker 2007, pp. 11–15). This is typically done by conducting so-called listening experiments under defined and reproducible conditions with several human listeners, the subjects. In the course of these experiments, stimulus properties are varied systematically and resulting hearing sensations are observed (Fastl and Zwicker 2007, p. 11). Psychoacoustics can be considered a field of research belonging to the more general class of hearing research, as for example also Audiology and medical acoustics. Hearing research aims at studying the hearing system, often requiring controllable stimulus presentation methods, especially including reproducible and physically well-defined audio playback procedures.

3.2.1 Sensations, Methodical Aspects, and Statistical Analysis

Hearing sensation[1] properties are quantified in Psychoacoustics observing or interviewing human subjects, not by physical measurements (Fastl and Zwicker 2007, pp. 11–15). For that reason, it is fundamental to distinguish between the actual hearing sensations and the subjects' judgments of the hearing sensation properties (Schneider and Parker 1990). In case a strict differentiation is not possible, the results have to be interpreted with regard to the possible deviation between hearing sensation properties and the ratings thereof.

A hearing sensation typically evolves as the result of multiple stimuli, in most cases perceived by different sensory modalities (Nathanail et al. 1996, Blauert and Jekosch 1997, Beerends and de Caluwe 1999, Fastl 2004, Menzel et al. 2008). Especially visual stimuli are known to influence and even dominate hearing sensations (Winkler 1992, Seeber 2002b, Völk et al. 2010b, Schmidhuber et al. 2011, Menzel 2011). The visual stimulation may be controlled during the course of a listening experiment by conducting the experiment in complete darkness. If not denoted otherwise, the listening experiments described in this thesis are carried out in complete darkness, assuming darkness as the visual reference stimulus, and the hearing sensations evolving in darkness as the reference sensations, which could possibly be altered by other visual stimulation. The influences of multiple stimuli or stimulus properties on a hearing sensation are typically addressed by varying only one stimulus property at a time.

An additional issue to be considered in Psychoacoustics is the cognitive influence on hearing sensations. Cognitive influence means that factors as for example the stimulus context (Schneider and Parker 1990), knowledge about the actual or a possible sound source (Fastl 2001), and selective attention that is focusing on a stimulus property (Lavie et al. 2004) are potentially taken into account in forming hearing sensations. For example, prior listening exposure to reverberant environments is shown by Zahorik et al. (2009) to increase the speech intelligibility at the cost of the ability to discriminate changes in the spatial properties of a sound field. Cognitive as well as learning effects due to prior listening exposure can be reduced by employing synthetic stimuli as for example pure tones or noises of different spectral shape, assumed not directly associated with a context or sound source and unlikely affected by selective attention. With the aim of reducing the meaning of natural or technical sounds, a procedure introduced by Fastl (2001) referred to as neutralizing can be applied, modifying a sound signal by spreading its spectral energy distribution while approximately preserving its loudness-time function. However, it is impossible to fully exclude cognitive effects from the hearing process, since they typically occur unconsciously and individually, affected by each subject's prior listening experience.

It is commonly attempted to reduce cognitive and individual influences on the results of listening experiments in general by including multiple subjects, a sample of subjects, and by using their results as the basis of an inter-individual statistical analysis. For small and medium sample sizes $N_\mathrm{S} \approx 10$ as frequently employed in Psychoacoustics, non-parametric methods are usually applied (Bortz 2005, Fastl and Zwicker 2007, pp. 11–15). Using this evaluation method, the inter-individual median of the individual results is regarded as the

[1] Blauert (1997, pp. 2–5) refers to acoustic stimuli as sound events and to hearing sensations as auditory events. Here, the evident and modality independent terms stimuli and sensations are used.

result of a typical subject, influenced only by cognitive effects and prior listening experience common to the sample as a whole. The associated inter-individual inter-quartile range of the individual results indicates the conformity of the sample and therefore the *variability of the typical subject's result.* In order to justify this procedure, the stimuli must be defined independently of the individual listener (cf. section 3.2.2). However, in conventional HP reproduction, inter-individually different stimuli occur, increasing the variability of the typical subject's result. This procedural aspect has to be considered when discussing the results of psychoacoustic experiments with conventional HP playback (cf. section 3.2.2). A quantitative assessment of the inter- and intra-individual variability of the transfer characteristics of three HP models frequently used in Psychoacoustics and hearing research in general is given in section 5.2 (cf. Völk 2011a).

Experimental results are frequently averaged not only inter- but also intra-individually, over some repetitions of the experiment, commonly employing non-parametric methods, due to the typically small number of repetitions $N_\mathrm{R} \approx 3$. This procedure increases the representativeness of the individual results, indicated by the intra-individual medians, and allows assessing the accuracy of the individual results by the intra-individual inter-quartile ranges. The inter-individual median of the intra-individual inter-quartile ranges represents the *accuracy of the typical subject.* The corresponding inter-individual inter-quartile range indicates the conformity of the sample regarding accuracy, the *variability of the accuracy.*

One- or multi-factorial analysis of variance (ANOVA) with complete repetition of measurement of the intra-individually averaged results indicates according to Bortz (2005) the statistical significance of the results. Throughout this work, effects are considered significant respectively highly significant, if the level of significance lies at or below 5% (indicated by $*$) respectively 1% ($**$). The resulting F- and p-values are given alongside the data. In multi-factorial configurations, interactions between factors are indicated by ×. Post-hoc multiple comparisons using Scheffé's method (Bortz 2005) are employed to identify significant and highly significant differences between factor levels.

3.2.2 Stimuli and Presentation Methods

In basic psychoacoustic studies, a frontally incident plane wave propagating in the free sound field is frequently desirable, for providing a simple while exactly defined stimulus (Zwicker and Feldtkeller 1967, p. 27). Assuming the ear is sensitive to sound pressure (Fastl and Zwicker 2007, p. 25), the undistorted pressure field in absence of the listener is often defined as the stimulus (Feldtkeller and Zwicker 1956, p. 6). Feldtkeller and Zwicker (1956, p. 1) even state that the *"plane sound field is the ideal sound field for all acoustic studies."* However, especially when examining spatial hearing, more complex scenarios, for example consisting of one or more point sources, must be considered in addition (Blauert 1997, pp. 27–30). In this case, the stimulus that is, according to Feldtkeller and Zwicker (1956, pp. 5–6), the spatio-temporal acoustic wave field becomes more complicated. In every case, the physical stimulus properties, in spatial hearing especially source position and dimensions, should be mathematically clearly describable, thus allowing for the definition of psychoacoustic relationships (cf. definition 17, Völk 2012b).

Definition 17 (*Stimuli for Traditional Psychoacoustic Experiments*)

The spatio-temporal acoustic pressure field eliciting hearing sensations during a listening experiment is, in absence of the listener, regarded as the stimulus for traditional psychoacoustic experiments not employing virtual acoustics or headphones.

There are two fundamentally different ways for presenting sounds in listening experiments: LS and HP playback (Fastl and Zwicker 2007, p. 5). Stimuli presented by LSs are fully covered by definition 17, while no spatially distributed sound field arises in absence of the listener in conventional HP reproduction. In that case, definition 17 does not apply, and an alternative stimulus definition is necessary. Two stimulus definitions are frequently employed in HP reproduction (definition 18): either the sound pressure in the auditory canal is defined as the stimulus, which is possible for monotic, diotic, and dichotic playback (Spikofski et al. 1986), or the reference field of free-field or diffuse-field equalization at a fixed sound pressure level is considered the stimulus, necessarily requiring diotic playback (Zwicker and Gässler 1952). The reference field level is typically measured at the head position of the equalization measurement, with the listener absent (cf. section 3.1.1).

Definition 18 (*Stimuli in Conventional Headphone Reproduction*)

The stimulus in conventional headphone reproduction is either the sound pressure in the auditory canal or the sound pressure level at the listening position with the listener absent in the reference sound field of free-field or diffuse-field equalization.

Defining the sound pressure in the auditory canal as the stimulus may be incorrect if the HP based listening experiment is designed to study effects of listening without HPs, since the sound pressure in the auditory canal is influenced by the individual listener's head and body, even in a free sound field, and is therefore different for each subject (Blauert 1997, pp. 78–93). If the results of a typical subject are to be found and a reference scene or natural listening situations in general are addressed, defining the auditory canal sound pressure as the stimulus is an improper choice since the corresponding reference scene signals are different for each subject. In such cases, an HP reproduction system is necessary, capable of reproducing the reference scene eardrum pressure signals individually. Such a system is provided by individual dynamic BS (Møller 1992, Völk 2011b), which can be regarded in the context of hearing research as an audio-playback procedure combining the advantages of HP and LS reproduction. The discussion of BS in the following chapters as the exemplary VA system evaluated with the methodical and procedural framework introduced covers this application of BS.

3.2.3 Virtual Acoustics Playback in Hearing Research

Regarding the audio playback in hearing research, the classes of approaches to VA defined in section 3.1 show different properties. Psychoacoustically motivated approaches are not appropriate for basic Psychoacoustics and hearing research since they are not aiming at generating a defined stimulus (definition 9). Physically motivated approaches on the contrary are by definition aiming at generating predefined physical conditions (cf. definition 14) and are therefore theoretically suited perfectly for the audio playback in Psychoacoustics and hearing research. However, regarding the correctness of the intended stimulus, it is crucial how well the goal of the physically motivated approach is fulfilled. It must be ensured that a playback method generates the intended stimulus before it is used to study psychoacoustic relationships and auditory perception in general. An exemplary validation of the physically motivated playback method BS for the audio playback in hearing research is given in the following chapters, including advantages and limitations.

3.2.4 Conventional Headphone Reproduction in Hearing Research

Traditionally, HPs were often preferred versus LSs for the audio playback in hearing research since their TFs are typically not altered remarkably by rooms with different acoustical conditions and HPs frequently employed for basic psychoacoustic research[2] produce significantly less nonlinear distortion than LSs (Zwicker and Gässler 1952, Fastl and Zwicker 2007, pp. 5–8). Furthermore, generating a plane wave using LSs requires high effort (Port 1964), while the free-field equalization frequently applied in HP reproduction (Fastl and Zwicker 2007, p. 8), allowing for playback at a defined free-field equivalent level (cf. section 3.1.1), can be implemented less expensively. However, free-field equalized HP reproduction belongs to the class of psychoacoustically motivated approaches to VA and is therefore not suited perfectly as a playback method for psychoacoustic experiments, especially if a physically well-defined stimulus is required. It is in general not *"possible to elicit the same hearing sensation as a plane sound-field by* [free-field equalized] *electrodynamical headphones"*, as erroneously stated by Feldtkeller and Zwicker (1956, p. 5). This becomes apparent looking at the hearing sensation positions corresponding to broad-band noise presented as a frontally incident plane wave and by free-field equalized HPs: the plane wave stimulus is perceived externalized as if stemming from a distant sound source (Völk 2010b), while free-field equalized HPs elicit a hearing sensation located in the head (Fastl and Zwicker 2007, p. 308).

However, the free-field equalized HP playback setup described by Fastl and Zwicker (2007, pp. 5–8) thoroughly defines a reproducible playback procedure for psychoacoustic experiments, offering the advantages of HP playback discussed above, namely less nonlinear distortion than LSs and independence of a specific listening environment. Further, the setup has proven valid and helpful in numerous basic scientific (Zwicker 1976b, Fastl 1976, Fastl and Bechly 1981, Zwicker 1984) and applied problems (Zwicker 1977, Zwicker and Dallmayr 1984, Zwicker 2000, Fastl 2000, Fastl et al. 2006, Fastl 2007, Fastl et al. 2009, DIN 45 631, DIN 45 631/A1, DIN ISO 226). Consequently, deviations of the results of listening experiments caused by erroneously assuming free-field hearing sensations to be present when using free-field equalized HP playback turned out to be of minor importance for many studies. In addition, the setup provides a well-defined and reproducible playback situation including an inter-individually averaged stimulus definition. Therefore, the setup can be regarded as a valid tool for basic psychoacoustic studies, as long as the circumstances discussed above are considered in evaluating the results. In this context,

[2] e. g. Beyer DT 48, Beyerdynamic DT 48 A, Sennheiser HD 650, Stax λ pro NEW

it is especially important to take into account that free-field equalized HPs as defined here allow for the playback at a free-field equivalent level only if the loudness comparison is carried out with diotic presentation and binaural listening. Conducting the loudness comparison monaurally results at frequencies below 4 kHz in more than 5 dB difference to the binaural situation (section 5.4.5 and Bocker and Mrass 1959). The difference between monaural and binaural loudness comparisons shows that the setup discussed here is not capable of providing free-field equivalent levels for monotic presentation, as erroneously assumed for example by Schorer (1986, 1988, 1989). Free-field equivalent levels are given for the free-field equalized HP playback setup proposed by Fastl and Zwicker (2007, p. 7) only for binaural listening to the diotic presentation.

Regardless of which presentation or calibration method is used for HP playback, the variability of the HP transfer characteristics defines how accurately stimuli can be applied. Since, in a typical psychoacoustic experiment, different subjects are included, inter-individual differences in the HP transfer characteristics have to be considered, especially if the calibration is done indirectly by means of the HP input voltage, using the HP sensitivity to compute the reference scene pressure or the corresponding level. In that case, inter-individual differences in the HP transfer characteristics result in different stimuli. Further, the intra-individual variability due to HP repositioning has to be taken into account, especially if listening experiments with repeated tracks are regarded. Variability results for three exemplary HP specimens are given in section 5.2 (cf. Völk 2011a).

3.2.5 Psychoacoustic Methods in Virtual Acoustics

It is of fundamental interest for most VA applications whether the signal processing of the VA system is audible in comparison to the reference scene for the majority of listeners, that is to say for the typical subject. Inaudible signal processing and the corresponding algorithms and systems are referred to as transparent procedures (cf. definition 19).

Definition 19 (*Transparent Signal Processing, Algorithms, and Systems*)

Audio signals are processed transparently if the processed signal is indistinguishable from the original for the majority of listeners. In that case, transparent audio signal processing algorithms and systems are employed.

For VA rendering, physically existing secondary sound sources are used with the aim of creating a sound field that would be generated by one or more physically non-existing virtual or primary sound sources (cf. definition 16), or of creating the corresponding perceptions. The signals fed to the secondary sound sources are referred to as driving signals, with corresponding driving functions in the frequency domain (definition 20, cf. Berkhout et al. 1993).

Definition 20 (*Driving Signals and Driving Functions*)

The secondary sources in virtual acoustics are fed by driving signals, which are in their frequency domain representation referred to as driving functions.

A psychoacoustic experiment addressing properties of VA systems is usually conducted by varying the virtual physical properties of the physically non-existing primary sources, controlled indirectly by the rendering algorithm, and by inspecting related changes of the hearing sensations. In other words, the primary source properties are regarded as the stimulus properties (Seeber 2002b, Menzel et al. 2006, Wittek et al. 2007b). This procedure is summarized by definition 21.

Definition 21 (*Stimuli for Psychoacoustic Studies in Virtual Acoustics*)

> *The primary spatio-temporal acoustic wave field that would elicit hearing sensations in the scenario to be simulated is regarded in absence of the listener as the stimulus for psychoacoustic experiments in virtual acoustics.*

This procedure is in line with traditional psychoacoustic experiments insofar as relations between variations in physical parameters and corresponding changes in hearing sensations are assessed. However, a major difference to traditional experiments is that the stimulus changes are caused indirectly by changing the rendering equations. This procedure results in loss of generality if the perfect synthesis is not achieved, since the specific VA setup characteristics, especially signal processing and hardware properties, are inevitably included in the results. Consequently, in addition to the traditional description of the experimental setup and procedure, details about the synthesis algorithm, its implementation, and the electroacoustic signal processing chain must be specified along with the results of psychoacoustic experiments in VA to allow for a meaningful discussion of the results.

3.3 Summary

Initially in this chapter, a classification scheme for approaches to virtual acoustics is proposed. Approaches are assigned, based on their methodical concept, to the classes of physically and psychoacoustically motivated procedures. Using the classification scheme, an overview of methods for the generation of virtual acoustical environments is given, motivating the selection of binaural synthesis as the exemplary method to be discussed in detail in the following chapters. Binaural synthesis as a physically motivated approach to virtual acoustics is considered suited best for the application in hearing research, while psychoacoustically motivated approaches are identified as principally not suited for that purpose. Furthermore, stimulus definitions for virtual acoustics, loudspeaker playback, and conventional headphone reproduction are given, allowing to correctly apply psychoacoustic methods to the quality evaluation of virtual acoustics systems, and to correctly employ conventional audio reproduction and virtual acoustics playback in hearing research. Thereby, a common lack of a thorough stimulus definition in conventional headphone reproduction independent of the applied equalization procedure is identified and traced back to the underlying psychoacoustically motivated concept. A reflection on aspects of employing psychoacoustic methods in virtual acoustics concludes the chapter.

4 Theoretical Aspects of Idealized Binaural Synthesis

This chapter deals with the system-theoretic background of binaural synthesis, not with implementation details, and is therefore based on an idealized best case situation. Prerequisites and assumptions necessary to derive the binaural synthesis theory are identified and marked, but not validated in the present chapter. The validation is given in chapter 5, along with the consequences arising when prerequisites or assumptions are not fulfilled. However, in allowing for the derivation of the theoretical and methodical basis of binaural synthesis, the best case scenario studied here represents the baseline situation for further considerations and is therefore discussed separately.

4.1 Main Principle

In a fundamental paper on binaural technology, Møller defined the basic idea of binaural recording in 1992 as follows:

> *"The input to the hearing consists of two signals: sound pressures at the eardrums. If these are recorded in the ears of a listener and reproduced exactly as they were, then the complete auditive experience is assumed to be reproduced, including timbre and spatial aspects."*

This binaural recording of the sound pressure signals at the eardrums, the so-called ear signals (cf. definition 22), is a time consuming process, especially if more than one listener, different head rotations, and varying listening positions are taken into account.

Definition 22 (*Ear Signals*)

The sound pressure signals detected by the eardrums are referred to as the ear signals.

Further, ear signal recording is possible only approximately (cf. section 5.1.4) and cognitive, memory, and inter-modal effects may influence the hearing sensations or the judgments thereof (cf. section 3.2.1). However, binaural recording has been in use for about half a century and still is applied frequently (early work e. g. by Bixler 1953, Wilkens 1972, Blauert et al. 1978, for overviews cf. Hammershøi and Møller 2002 or Daniel et al. 2007).

A less time consuming method of generating ear signals also attributed to the binaural technologies is binaural synthesis (BS), described for example by Wightman and Kistler (1989a,b) or Møller (1992). In contrast to binaural *recording*, in binaural *synthesis* the ear signals are generated by convolving the input signal of each single sound source with the impulse responses (IRs) of the paths between the source and the eardrums in the corresponding real life situation (e. g. Haferkorn and Schmid 1996). In general, these paths depend on the positions and orientations of the sound sources and the body as well as on the listening environment, and are therefore time *variant.* However, by an IR, a system is modeled linear and time *invariant* (Oppenheim et al. 1998). According to Terhardt (1998, p. 61), linearity is usually given for acoustic systems. Time-variant systems can be approximated by series of M_{ir} IRs, assuming the system time invariant within temporal intervals τ_{i}, with $i = 0, \ldots, M_{\mathrm{ir}} - 1$ (cf. section 5.1.1). To update the simulation, the IR representing the present system state is selected by an adaptive signal processing routine. The accuracy of the procedure depends on M_{ir}, the duration of the temporal intervals τ_{i}, the implementation of the signal processing, and the nature of the time-invariant system.

Consequently, it is possible to approximately describe each of the systems involved in a real acoustical scenario by its IR $h(t)$ or by an array of IRs $h_{\tau_{\mathrm{i}}}(t)$. The IRs for BS are according to definition 23 referred to as binaural impulse response pairs (BIRPs), with corresponding binaural transfer function pairs (BTFPs).

Definition 23 (*Binaural Impulse Response and Transfer Function Pairs*)

The impulse responses of the transfer paths from a sound source input signal to the ear signals are referred to as binaural impulse response pairs with corresponding binaural transfer function pairs.

BIRP updates are triggered in this so-called dynamic BS in contrast to static procedures based on listening environment data as well as body and sound source positions and orientations. Updates occur if the listener moves, if the source is repositioned, or if other parts of the situation change (Foster 1992, Mackensen et al. 1999). Complex acoustic situations with more than one source are generated by linear superposition of the ear signals evolving from the single sources. If the BIRPs are acquired by measurement, the BS is referred to as data based, in contrast to model based approaches (definition 24). In the model based case, the BIRPs are synthesized by Auralization (Vorländer 2008).

Definition 24 (*Data or Model Based Static or Dynamic Binaural Synthesis*)

Binaural synthesis is an audio reproduction procedure aiming at generating the ear signals of a real or hypothetical reference scene by convolving the binaural impulse response pairs of the reference scene with the source signals. If the impulse responses are acquired by measurement, the synthesis is referred to as data based, otherwise as model based. In the case of dynamic binaural synthesis, in contrast to static binaural synthesis, the impulse responses are adapted while the system is in operation.

By now, a variety of BS systems is available and in use, for example in virtual and augmented reality (Begault 1999, Gröhn et al. 2007, Völk et al. 2007) or psychoacoustic research (Blauert et al. 2000, Djelani et al. 2000, Zahorik 2002, Völk 2009). If certain prerequisites are met and corrections are applied, presenting the convolution products of a BS system via headphones (HPs) can generate the reference scene ear signals with a degree of authenticity depending on the accuracy of the BS procedure. In the following, the necessary restrictions and corrections are identified and discussed step by step by a system-theoretic analysis of all components involved. Based on the requirements, different equalization procedures are presented with respect to the achievable overall system transfer functions (TFs). The system-theoretic framework derived allows for mathematically predicting the accuracy of the ear signals generated by a specific BS approach. On this basis, identifying the procedure suited best for a specific application is possible, including requirements and restrictions (cf. Völk 2011b).

While derived based on a static situation, this chapter applies to static and dynamic BS. Specific aspects of dynamic systems are discussed in section 5.1. Further, the theory discussed by the example of data based synthesis also applies to model based approaches. In the model based case, the results of the different measurements described in the following must be regarded as the auralization targets. Furthermore, the system-theoretic framework inherently covers binaural recording, which is not discussed separately, since it represents a special case of the more general concept of static BS.

Table 4.1 gives an overview of the structure of this chapter. It is intended to derive a system-theoretic description of the BS of a reference scene by discussing three major scenarios (rows of table 4.1). Each scenario consists of a sound source, driven in all scenarios by the same input signal (left column), causing via different transfer paths (center column) different ear signals (right column). The symbols used in the table are introduced in the respective sections and tabulated in appendix B.

Source signal sequence s_{ls}	*Reference scene* $\mathbf{h}_{\text{ref}}^{\text{ind}}$ (section 4.2)	*Ear signals* $\mathbf{p}_{\text{e}}^{\text{ind}}$
	Non-equalized binaural synthesis $\mathbf{h}_{\text{ne}}^{\text{ind,h}}$ (section 4.5) Recording (section 4.3) Playback (section 4.4)	*Ear signals* $\mathbf{p}_{\text{e}_{\text{ne}}}^{\text{ind,h}}$
	Equalized binaural synthesis $\mathbf{h}_{\text{bs}}^{\text{ind,h}}$ (section 4.7) Recording (section 4.3) Equalization (section 4.6) Playback (section 4.4)	*Ear signals* $\mathbf{p}_{\text{e}_{\text{bs}}}^{\text{ind,h}}$

Table 4.1: Schematic overview of the paths between the source and ear signals used for the system-theoretic derivation of binaural synthesis. Depicted are signals and impulse responses including the nomenclature and references to the corresponding sections.

In section 4.2 (first row of table 4.1), the *reference scene* to be simulated is defined and described system theoretically, including two approaches of approximately measuring the reference scene ear signals: probe microphone and artificial head (AH) recording. On this basis, three ways of implementing the BIRP *recording situation* are introduced in section 4.3 (second row): probe microphone recording with the probe tube tips in the auditory canals close to the eardrums, miniature microphone recording at the entrances to the blocked auditory canals, and conventional AH recording with the microphones at the eardrum positions. In section 4.4, the *HP playback situation* is defined, including the playback paths of the actual synthesis situation and their approximation by the three measurement methods of the recording section. The approximations are typically used for equalizing a BS system. All combinations of recording, equalization, and playback situations are formulated in section 4.6 (third row), based on the non-equalized combinations of the recording and playback situations given in section 4.5. Summarizing, the results of all equalized BS procedures are presented and discussed in section 4.7.

4.2 Reference Scene

The eventual aim of BS, as a virtual acoustics procedure, is according to definition 1 to reconstruct the hearing sensations occurring in a reference scene, for example a loudspeaker (LS) in a reverberant room. This situation is regarded as the reference scene for the derivation of the BS theory in this chapter.

Definition 25 (*Reference Scene for Binaural Synthesis*)

A subject listening to a loudspeaker in a reverberant room represents the reference scene for deriving the binaural synthesis theory. This situation is assumed static.

Signals occurring in the reference scene according to definition 25 are the time varying ear signals $p_{\mathrm{e_L}}$ and $p_{\mathrm{e_R}}$, the LS input voltage u_{ls}, and the digital sample sequence s_{ls} encoding the signal to be played back by the LS. Regarding the synthesis of single sound sources, the restriction to an LS does not result in a loss of generality, since the LS can be assumed an ideal electroacoustic transducer (e. g. a monopole point source, Zollner and Zwicker 1993, pp. 75–77). The synthesis of multiple independent sources is possible by linearly superposing the respective ear signals. In general, linear interactions between sources, as for example mutual reflections, are simulated correctly if the IRs are measured in the system state to be simulated (Völk et al. 2010a). Mutual interactions modifying system characteristics during simultaneous operation, as for example mutual radiation impedance effects of closely spaced LSs, are not covered by the linear time-invariant system theory applied. However, these effects are negligible in most situations (Völk et al. 2010a).

The BS theory described is limited to sources whose TFs can be measured or simulated. In other cases, it may be helpful to apply binaural recording or to approximate the source characteristics combining feasible basis functions (e. g. Wefers et al. 2011).

The ear signals in the reference scene according to definition 25 (subscript *ref*) include influences of systems grouped by their affiliation to the generation or propagation parts of the sound transfer paths from the source input to the eardrums. The generation part of both paths is the audio output system defined by equation 2.7, while the propagation parts are defined between u_{ls} and the ear signals $\mathbf{p}_{\mathrm{e}}^{\mathrm{ind}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}})$ by the BTFPs

$$\mathbf{H}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{e}}}^{\mathrm{ind}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}) = \frac{\mathbf{P}_{\mathrm{e}}^{\mathrm{ind}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}})}{U_{\mathrm{ls}}}. \tag{4.1}$$

The BTFPs contain the transfer characteristics of the LS and the sound propagation paths including all reflection and diffraction effects as for example room reflections and transformation characteristics of the outer ears and other body parts. Consequently, the BTFPs depend on the position and orientation of the listener and are therefore modeled by an array of TF pairs, with the subject and LS position vectors as the array parameters.

The BTFPs defined by equation 4.1 are by some authors referred to as head-related transfer functions (HRTFs), if a free-field reference scene is discussed (e. g. Wightman and Kistler 2005). However, HRTFs are commonly defined as the relations of the eardrum sound pressure spectra $\mathbf{P}_{\mathrm{e}}^{\mathrm{ind}}(\mathbf{x}_{\mathrm{h}}, \mathbf{x}_{\mathrm{ls}})$, measured under free-field conditions, to the sound pressure spectrum at the midpoint of the recording head, measured in the same situation but in the absence of the head (Blauert 1997, p. 78, free-field transfer function). The corresponding TFs under reverberant conditions are according to Møller (1992) referred to as binaural room transfer functions (BRTFs). The labels and concepts of HRTFs and BRTFs are avoided here due to practical difficulties in recording the sound pressure at a single point without errors introduced by the measurement system, especially under reverberant conditions (Zollner 1982, 1995, Zollner and Zwicker 1993, p. 182). In this work, TFs are defined and referred to mathematically. Where general denominators are necessary, the neutral labels BTFPs respectively BIRPs are used (cf. definition 23).

Combining equation 4.1 and the audio output system description given by equation 2.7, the reference scene ear signal spectra

$$\mathbf{P}_{\mathrm{e}}^{\mathrm{ind}}(\mathbf{x}_{\mathrm{h}_{\mathrm{ref}}}, \mathbf{x}_{\mathrm{ls}_{\mathrm{ref}}}) = S_{\mathrm{ls}} \cdot H_{\mathrm{o}_{\mathrm{ref}}} \cdot \mathbf{H}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{e}}}^{\mathrm{ind}}(\mathbf{x}_{\mathrm{h}_{\mathrm{ref}}}, \mathbf{x}_{\mathrm{ls}_{\mathrm{ref}}}) \tag{4.2}$$

are formulated dependent on the spectrum of the LS driving sequence. Using equation 4.2, the reference scene TFs connecting the spectrum of the LS driving sequence to the reference scene ear signal spectra are defined by

$$\mathbf{H}_{\mathrm{ref}}^{\mathrm{ind}}(\mathbf{x}_{\mathrm{h}_{\mathrm{ref}}}, \mathbf{x}_{\mathrm{ls}_{\mathrm{ref}}}) = \frac{\mathbf{P}_{\mathrm{e}}^{\mathrm{ind}}(\mathbf{x}_{\mathrm{h}_{\mathrm{ref}}}, \mathbf{x}_{\mathrm{ls}_{\mathrm{ref}}})}{S_{\mathrm{ls}}} = H_{\mathrm{o}_{\mathrm{ref}}} \cdot \mathbf{H}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{e}}}^{\mathrm{ind}}(\mathbf{x}_{\mathrm{h}_{\mathrm{ref}}}, \mathbf{x}_{\mathrm{ls}_{\mathrm{ref}}}). \tag{4.3}$$

To reflect the actual hearing process, ear signal measurements must capture the pressure distribution across the eardrum, which is possible only approximately (cf. section 5.1.4). In the following, two methods of approximating the reference scene TFs given by equation 4.3 are discussed: probe microphone measurement with the probe tube tips in the auditory canals of an individual or artificial head, close to the eardrums, and conventional AH measurement with the microphones at the eardrum positions. The sound pressure signals detected by both methods deviate from the actual reference scene ear signals. Probe microphone measurements are valid only for the probe tube tip positions and suffer from methodical difficulties (cf. section 5.1.4). Further, individual or artificial head measurements deviate from the reference scene due to inter-individual differences and modeling errors, as for example missing hair or torso (Katz 2000, Minnaar et al. 2001).

The miniature microphone measurement at the blocked auditory canal entrance is not a procedure of approximating ear signals, and is therefore not included in the present section. However, blocked auditory canal entrance measurements are discussed in section 4.3.2 as a method of implementing the BS recording situation.

Probe Microphone Reference Scene Approximation Probe microphone (subscript *pm*) approximations are valid for the probe microphone tube tip positions $\mathbf{x}_{\mathrm{pm}}$, which are aiming at ear signal measurements typically selected close to the eardrums. The reference scene propagation paths from the LS input voltage to the sound pressures at the probe tube tips are described by the TFs

$$\mathbf{H}^{\mathrm{ind}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{pm}}}(\mathbf{x}_{\mathrm{h}_{\mathrm{ref}}},\mathbf{x}_{\mathrm{ls}_{\mathrm{ref}}},\mathbf{x}_{\mathrm{pm}_{\mathrm{ref}}}) = \frac{\mathbf{P}^{\mathrm{ind}}_{\mathrm{pm}}(\mathbf{x}_{\mathrm{h}_{\mathrm{ref}}},\mathbf{x}_{\mathrm{ls}_{\mathrm{ref}}},\mathbf{x}_{\mathrm{pm}_{\mathrm{ref}}})}{U_{\mathrm{ls}}}. \tag{4.4}$$

The reference scene propagation path TFs, connecting the LS input voltages with the probe microphone output voltages, are defined in a similar way by

$$\mathbf{H}^{\mathrm{ind}}_{u_{\mathrm{ls}},\mathbf{u}_{\mathrm{pm}}}(\mathbf{x}_{\mathrm{h}_{\mathrm{ref}}},\mathbf{x}_{\mathrm{ls}_{\mathrm{ref}}},\mathbf{x}_{\mathrm{pm}_{\mathrm{ref}}}) = \frac{\mathbf{U}^{\mathrm{ind}}_{\mathrm{pm}}(\mathbf{x}_{\mathrm{h}_{\mathrm{ref}}},\mathbf{x}_{\mathrm{ls}_{\mathrm{ref}}},\mathbf{x}_{\mathrm{pm}_{\mathrm{ref}}})}{U_{\mathrm{ls}}}. \tag{4.5}$$

Relating the sound pressure spectra at the probe microphones to the spectrum of the LS driving sequence results in the TFs

$$\mathbf{H}^{\mathrm{ind}}_{\mathrm{ref}_{\mathrm{pm}}}(\mathbf{x}_{\mathrm{h}_{\mathrm{ref}}},\mathbf{x}_{\mathrm{ls}_{\mathrm{ref}}},\mathbf{x}_{\mathrm{pm}_{\mathrm{ref}}}) = \frac{\mathbf{P}^{\mathrm{ind}}_{\mathrm{pm}}(\mathbf{x}_{\mathrm{h}_{\mathrm{ref}}},\mathbf{x}_{\mathrm{ls}_{\mathrm{ref}}},\mathbf{x}_{\mathrm{pm}_{\mathrm{ref}}})}{S_{\mathrm{ls}}}. \tag{4.6}$$

Consequently, the reference scene TFs defined by equation 4.3 are approximated using probe microphones with associated input systems ($\mathbf{H}_{\mathrm{pm}}$ and $\mathbf{H}_{\mathrm{i}_{\mathrm{pm}}}$, equation 2.6) by

$$\begin{aligned}\mathbf{H}^{\mathrm{ind}}_{\mathrm{ref}}(\mathbf{x}_{\mathrm{h}_{\mathrm{ref}}},\mathbf{x}_{\mathrm{ls}_{\mathrm{ref}}}) &\approx \mathbf{H}^{\mathrm{ind}}_{\mathrm{ref}_{\mathrm{pm}}}(\mathbf{x}_{\mathrm{h}_{\mathrm{ref}}},\mathbf{x}_{\mathrm{ls}_{\mathrm{ref}}},\mathbf{x}_{\mathrm{pm}_{\mathrm{ref}}})\\ &= H_{\mathrm{o}_{\mathrm{ref}}} \cdot \mathbf{H}^{\mathrm{ind}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{pm}}}(\mathbf{x}_{\mathrm{h}_{\mathrm{ref}}},\mathbf{x}_{\mathrm{ls}_{\mathrm{ref}}},\mathbf{x}_{\mathrm{pm}_{\mathrm{ref}}}) = \frac{\mathbf{S}^{\mathrm{ind}}_{\mathrm{pm}}(\mathbf{x}_{\mathrm{h}_{\mathrm{ref}}},\mathbf{x}_{\mathrm{ls}_{\mathrm{ref}}},\mathbf{x}_{\mathrm{pm}_{\mathrm{ref}}})}{S_{\mathrm{ls}} \cdot \mathbf{H}_{\mathrm{pm}} \cdot \mathbf{H}_{\mathrm{i}_{\mathrm{pm}}}}.\end{aligned} \tag{4.7}$$

Artificial Head Reference Scene Approximation Ear signal approximations are also possible using AHs with microphones at the eardrum positions (superscript *ah*). The propagation paths between the LS input voltage and the sound pressures at the AH microphones (subscript *ahm*) are defined by the TFs

$$\mathbf{H}^{\mathrm{ah}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{h}_{\mathrm{ref}}},\mathbf{x}_{\mathrm{ls}_{\mathrm{ref}}}) = \frac{\mathbf{P}^{\mathrm{ah}}_{\mathrm{ahm}}(\mathbf{x}_{\mathrm{h}_{\mathrm{ref}}},\mathbf{x}_{\mathrm{ls}_{\mathrm{ref}}})}{U_{\mathrm{ls}}}. \tag{4.8}$$

The relation of the LS input voltage spectrum to the microphone output voltage spectra in the same situation is given by

$$\mathbf{H}^{\mathrm{ah}}_{u_{\mathrm{ls}},\mathbf{u}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{h}_{\mathrm{ref}}},\mathbf{x}_{\mathrm{ls}_{\mathrm{ref}}}) = \frac{\mathbf{U}^{\mathrm{ah}}_{\mathrm{ahm}}(\mathbf{x}_{\mathrm{h}_{\mathrm{ref}}},\mathbf{x}_{\mathrm{ls}_{\mathrm{ref}}})}{U_{\mathrm{ls}}}. \tag{4.9}$$

Further, relating the AH microphone pressure spectra and the LS driving spectrum by

$$\mathbf{H}^{\mathrm{ah}}_{\mathrm{ref}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{h}_{\mathrm{ref}}},\mathbf{x}_{\mathrm{ls}_{\mathrm{ref}}}) = \frac{\mathbf{P}^{\mathrm{ah}}_{\mathrm{ahm}}(\mathbf{x}_{\mathrm{h}_{\mathrm{ref}}},\mathbf{x}_{\mathrm{ls}_{\mathrm{ref}}})}{S_{\mathrm{ls}}} = H_{\mathrm{o}_{\mathrm{ref}}} \cdot \mathbf{H}^{\mathrm{ah}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{h}_{\mathrm{ref}}},\mathbf{x}_{\mathrm{ls}_{\mathrm{ref}}}) \tag{4.10}$$

allows in combination with equation 2.6 for approximating the reference scene TFs, mathematically described by

$$\mathbf{H}_{\mathrm{ref}}^{\mathrm{ind}}(\mathbf{x}_{\mathrm{h}_{\mathrm{ref}}}, \mathbf{x}_{\mathrm{ls}_{\mathrm{ref}}}) \approx \mathbf{H}_{\mathrm{ref}_{\mathrm{ahm}}}^{\mathrm{ah}}(\mathbf{x}_{\mathrm{h}_{\mathrm{ref}}}, \mathbf{x}_{\mathrm{ls}_{\mathrm{ref}}}) = \frac{\mathbf{S}_{\mathrm{ahm}}^{\mathrm{ah}}(\mathbf{x}_{\mathrm{h}_{\mathrm{ref}}}, \mathbf{x}_{\mathrm{ls}_{\mathrm{ref}}})}{S_{\mathrm{ls}} \cdot \mathbf{H}_{\mathrm{i}_{\mathrm{ah}}} \cdot \mathbf{H}_{\mathrm{ahm}}}. \tag{4.11}$$

The TFs defined in the present section describe the reference scene to be simulated, allowing for evaluating BS by comparing the reference and synthesis scene TFs.

4.3 Recording Situation

For the BIRP recording, a measurement signal is played back by the reference scene LS and recorded with an AH (Mackensen et al. 1999, Spikofski and Fruhmann 2001, Pellegrini et al. 2007) or probe microphones with the tube tips in the auditory canals (Wightman and Kistler 1989a,b, 2005). It is also common practice to use miniature microphones at the entrances to the blocked auditory canals (Møller et al. 1995b, Hammershøi and Møller 2002). According to Hammershøi and Møller (1996a,b), when recording at the blocked auditory canal entrances, all directional information is captured, but little individual characteristics. This can reduce errors if the recordings from one subject are used to synthesize ear signals for other subjects (nonindividual in contrast to individual recording, cf. section 5.2.4). Additional advantages are reduced complexity and increased signal to noise ratio compared to probe microphone recording (section 5.1.4). Frequency independent microphone TFs are of minor importance for BS as described here, since the microphone TFs are equalized in the BS process (section 4.6).

Regardless of the recording procedure, the resulting signals contain the transfer characteristics of the input and output systems and of the sound propagation paths between the LS input and microphone output voltages. The recording situation (subscript *rec*) is consequently described in general by the TFs

$$\begin{aligned}\mathbf{H}_{\mathrm{rec}_{\mathrm{mic}}}^{\mathrm{ind}}(\mathbf{x}_{\mathrm{h}_{\mathrm{rec}}}, \mathbf{x}_{\mathrm{ls}_{\mathrm{rec}}}, \mathbf{x}_{\mathrm{mic}_{\mathrm{rec}}}) &= H_{\mathrm{o}_{\mathrm{rec}}} \cdot \mathbf{H}_{u_{\mathrm{ls}},\mathbf{u}_{\mathrm{mic}}}^{\mathrm{ind}}(\mathbf{x}_{\mathrm{h}_{\mathrm{rec}}}, \mathbf{x}_{\mathrm{ls}_{\mathrm{rec}}}, \mathbf{x}_{\mathrm{mic}_{\mathrm{rec}}}) \cdot \mathbf{H}_{\mathrm{i}_{\mathrm{rec}}} \\ &= \frac{\mathbf{S}_{\mathrm{mic}}^{\mathrm{ind}}(\mathbf{x}_{\mathrm{h}_{\mathrm{rec}}}, \mathbf{x}_{\mathrm{ls}_{\mathrm{rec}}}, \mathbf{x}_{\mathrm{mic}_{\mathrm{rec}}})}{S_{\mathrm{ls}}}.\end{aligned} \tag{4.12}$$

Since the synthesis of the reference scene is intended, identical equipment and geometric relations are assumed in the recording situation and the reference scene (assumption 1).

Assumption 1 (*Output Equipment, Loudspeaker, Scenario*)

> *The reference scene output equipment and loudspeaker are used for the recording, with the head and loudspeaker positions identical to the reference scene.*

In the following paragraphs, all three recording procedures are described by means of the respective recording situation TFs. Alongside, the recording situation TFs are related to the reference scene TFs for each method, pointing out possibly occurring deviations.

4.3.1 Probe Microphone Recording

Probe microphone recording with the probe tube tips close to human or artificial head eardrums is employed in BS for example by Wightman and Kistler (1989a). This procedure is described with equations 2.6 and 4.12 by the TFs

$$\begin{aligned}\mathbf{H}^{\mathrm{ind}}_{\mathrm{rec}_{\mathrm{pm}}}(\mathbf{x}_{\mathrm{h}_{\mathrm{rec}}},\mathbf{x}_{\mathrm{ls}_{\mathrm{rec}}},\mathbf{x}_{\mathrm{pm}_{\mathrm{rec}}}) &= H_{\mathrm{o}_{\mathrm{rec}}}\cdot\mathbf{H}^{\mathrm{ind}}_{u_{\mathrm{ls}},\mathbf{u}_{\mathrm{pm}}}(\mathbf{x}_{\mathrm{h}_{\mathrm{rec}}},\mathbf{x}_{\mathrm{ls}_{\mathrm{rec}}},\mathbf{x}_{\mathrm{pm}_{\mathrm{rec}}})\cdot\mathbf{H}_{\mathrm{i}_{\mathrm{pm}}}\\ &= H_{\mathrm{o}_{\mathrm{rec}}}\cdot\mathbf{H}^{\mathrm{ind}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{pm}}}(\mathbf{x}_{\mathrm{h}_{\mathrm{rec}}},\mathbf{x}_{\mathrm{ls}_{\mathrm{rec}}},\mathbf{x}_{\mathrm{pm}_{\mathrm{rec}}})\cdot\mathbf{H}_{\mathrm{pm}}\cdot\mathbf{H}_{\mathrm{i}_{\mathrm{pm}}}\,.\end{aligned} \tag{4.13}$$

These recording situation TFs are related to the reference scene TFs of equation 4.7 by

$$\mathbf{H}^{\mathrm{ind}}_{\mathrm{rec}_{\mathrm{pm}}}(\mathbf{x}_{\mathrm{h}_{\mathrm{ref}}},\mathbf{x}_{\mathrm{ls}_{\mathrm{ref}}},\mathbf{x}_{\mathrm{pm}_{\mathrm{rec}}}) \approx \mathbf{H}^{\mathrm{ind}}_{\mathrm{ref}}(\mathbf{x}_{\mathrm{h}_{\mathrm{ref}}},\mathbf{x}_{\mathrm{ls}_{\mathrm{ref}}})\cdot\mathbf{H}_{\mathrm{pm}}\cdot\mathbf{H}_{\mathrm{i}_{\mathrm{pm}}}\,. \tag{4.14}$$

4.3.2 Blocked Auditory Canal Recording

Miniature microphone (subscript m) recording at the entrances to the blocked auditory canals (superscript b) is formulated using equations 2.6 and 4.12 by the TFs

$$\begin{aligned}\mathbf{H}^{\mathrm{ind,b}}_{\mathrm{rec}_{\mathrm{m}}}(\mathbf{x}_{\mathrm{h}_{\mathrm{rec}}},\mathbf{x}_{\mathrm{ls}_{\mathrm{rec}}},\mathbf{x}_{\mathrm{m}_{\mathrm{rec}}}) &= H_{\mathrm{o}_{\mathrm{rec}}}\cdot\mathbf{H}^{\mathrm{ind,b}}_{u_{\mathrm{ls}},\mathbf{u}_{\mathrm{m}}}(\mathbf{x}_{\mathrm{h}_{\mathrm{rec}}},\mathbf{x}_{\mathrm{ls}_{\mathrm{rec}}},\mathbf{x}_{\mathrm{m}_{\mathrm{rec}}})\cdot\mathbf{H}_{\mathrm{i}_{\mathrm{m}}}\\ &= H_{\mathrm{o}_{\mathrm{rec}}}\cdot\mathbf{H}^{\mathrm{ind,b}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{m}}}(\mathbf{x}_{\mathrm{h}_{\mathrm{rec}}},\mathbf{x}_{\mathrm{ls}_{\mathrm{rec}}},\mathbf{x}_{\mathrm{m}_{\mathrm{rec}}})\cdot\mathbf{H}_{\mathrm{m}}\cdot\mathbf{H}_{\mathrm{i}_{\mathrm{m}}}\,.\end{aligned} \tag{4.15}$$

These TFs are used in BS for example by Møller (1992), Hammershøi and Møller (2002), and Völk and Fastl (2011a). Equation 4.15 differs from the reference scene TF description in equation 4.3 even if the same equipment and geometric setup is used, formulated by

$$\begin{aligned}&\mathbf{H}^{\mathrm{ind,b}}_{\mathrm{rec}_{\mathrm{m}}}(\mathbf{x}_{\mathrm{h}_{\mathrm{ref}}},\mathbf{x}_{\mathrm{ls}_{\mathrm{ref}}},\mathbf{x}_{\mathrm{m}_{\mathrm{rec}}}) =\\ &\quad= \mathbf{H}^{\mathrm{ind}}_{\mathrm{ref}}(\mathbf{x}_{\mathrm{h}_{\mathrm{ref}}},\mathbf{x}_{\mathrm{ls}_{\mathrm{ref}}})\cdot\frac{\mathbf{H}^{\mathrm{ind,b}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{m}}}(\mathbf{x}_{\mathrm{h}_{\mathrm{ref}}},\mathbf{x}_{\mathrm{ls}_{\mathrm{ref}}},\mathbf{x}_{\mathrm{m}_{\mathrm{rec}}})}{\mathbf{H}^{\mathrm{ind}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{e}}}(\mathbf{x}_{\mathrm{h}_{\mathrm{ref}}},\mathbf{x}_{\mathrm{ls}_{\mathrm{ref}}})}\cdot\mathbf{H}_{\mathrm{m}}\cdot\mathbf{H}_{\mathrm{i}_{\mathrm{m}}}\,.\end{aligned} \tag{4.16}$$

TFs are valid system descriptions if the system is constant (section 2.2). The system described by equation 4.16 may change, for example if HPs are put on. In this case, the radiation impedance changes compared to the situation without the HPs (Møller et al. 1995a, Vorländer 2000, Völk 2010a).

Blocked auditory canal measurements are possible with human subjects and AHs (cf. section 5.3). The AH situation is covered by the framework in that the individual is replaced by an artificial head. If the microphones are positioned reproducibly at the entrances to the blocked AH auditory canals, the situation becomes independent of the microphone positions, which is typically not possible for human heads (cf. section 5.2.2). Figure 4.1 shows the auditory-adapted analysis (AAA) spectrogram defined in section 2.4 of a blocked auditory canal AH recording situation TF according to equation 4.16. Depicted is a typical BS reference scene[1] in a reverberant environment with the LS distance $r_{\mathrm{ls}} = 2\,\mathrm{m}$. In this case, the AH may be regarded as a specific individual head.

[1] Klein + Hummel Studio Monitor Loudspeaker O 98, custom-made artificial head $\mathrm{AH_c}$

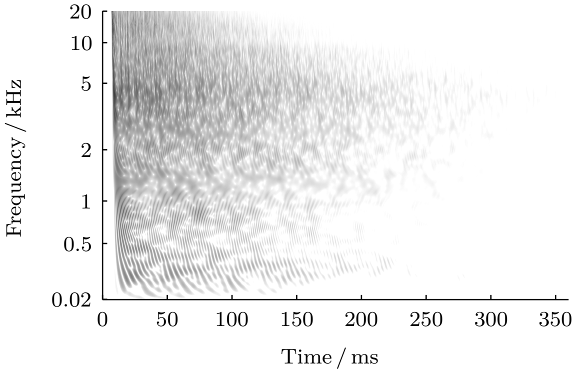

Figure 4.1: Auditory-adapted analysis spectrogram with 60 dB visible dynamic of the transfer path between a loudspeaker at 2 m distance in a reverberant environment and a microphone at the entrance to the blocked auditory canal of an artificial head. The path is defined between the digital sequences corresponding to the loudspeaker input and microphone output voltages.

The measurement depicted by figure 4.1 is independent of the microphone position since the AH used allows for reproducibly positioning the microphone at the blocked auditory canal entrance. Therefore, typical reference scene and recording situation parameters can be discussed and compared based on the data. From an auditory perspective, the transfer path from the LS input to the microphone output ports is with 60 dB decay times of up to more than 300 ms spectro-temporally effective (cf. definition 4). The frequency independent initial delay reflects the sound travel time $r_{\mathrm{ls}}/c \approx 5.8\,\mathrm{ms}$ from the LS to the microphone. Magnitude and phase show a rather frequency independent region between the lower limit of the LS transmission bandwidth at approximately 100 Hz and about 3 kHz, while the spectral interference pattern at higher frequencies is characteristic for the AH and its orientation with respect to the LS (Blauert 1997, pp. 88–92).

4.3.3 Artificial Head Recording

Using equations 4.12 and 2.6, the conventional artificial head recording situation with the microphones at the eardrum positions is described by the TFs

$$\mathbf{H}^{\mathrm{ah}}_{\mathrm{rec}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{h}_{\mathrm{rec}}}, \mathbf{x}_{\mathrm{ls}_{\mathrm{rec}}}) = H_{\mathrm{o}_{\mathrm{rec}}} \cdot \mathbf{H}^{\mathrm{ah}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{h}_{\mathrm{rec}}}, \mathbf{x}_{\mathrm{ls}_{\mathrm{rec}}}) \cdot \mathbf{H}_{\mathrm{ahm}} \cdot \mathbf{H}_{\mathrm{i}_{\mathrm{ah}}}\,. \tag{4.17}$$

These TFs are independent of the microphone positions since the AH microphones are fixed. A comparison to the AH reference scene TFs according to equation 4.11 reveals

$$\mathbf{H}^{\mathrm{ah}}_{\mathrm{rec}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{h}_{\mathrm{ref}}}, \mathbf{x}_{\mathrm{ls}_{\mathrm{ref}}}) = \mathbf{H}^{\mathrm{ah}}_{\mathrm{ref}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{h}_{\mathrm{ref}}}, \mathbf{x}_{\mathrm{ls}_{\mathrm{ref}}}) \cdot \mathbf{H}_{\mathrm{ahm}} \cdot \mathbf{H}_{\mathrm{i}_{\mathrm{ah}}}\,. \tag{4.18}$$

The AH recording TFs described by equation 4.17 differ from the individual reference scene TFs given by equation 4.3 assuming identical equipment and the same situation if the AH differs from the individual head, formulated by

$$\mathbf{H}^{\mathrm{ah}}_{\mathrm{rec}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{h}_{\mathrm{ref}}}, \mathbf{x}_{\mathrm{ls}_{\mathrm{ref}}}) = \mathbf{H}^{\mathrm{ind}}_{\mathrm{ref}}(\mathbf{x}_{\mathrm{h}_{\mathrm{ref}}}, \mathbf{x}_{\mathrm{ls}_{\mathrm{ref}}}) \cdot \frac{\mathbf{H}^{\mathrm{ah}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{h}_{\mathrm{ref}}}, \mathbf{x}_{\mathrm{ls}_{\mathrm{ref}}})}{\mathbf{H}^{\mathrm{ind}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{e}}}(\mathbf{x}_{\mathrm{h}_{\mathrm{ref}}}, \mathbf{x}_{\mathrm{ls}_{\mathrm{ref}}})} \cdot \mathbf{H}_{\mathrm{ahm}} \cdot \mathbf{H}_{\mathrm{i}_{\mathrm{ah}}}\,. \tag{4.19}$$

Figure 4.2 shows the AAA spectrogram of an example for the TFs defined by equation 4.19. The recording was carried out in the situation and with the AH of figure 4.1.

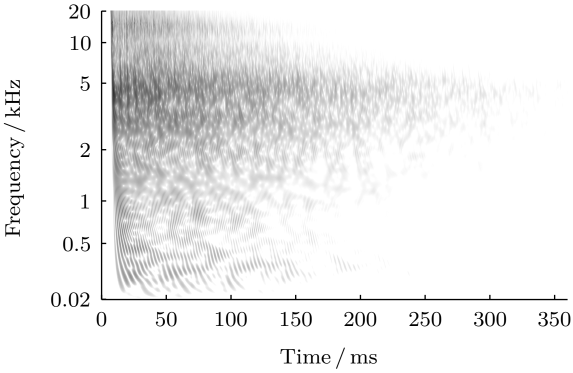

Figure 4.2: Auditory-adapted analysis spectrogram with 60 dB visible dynamic of the transfer path between a loudspeaker at 2 m distance in a reverberant environment and an artificial head microphone at the eardrum position. The path is defined between the digital sequences corresponding to the loudspeaker input and the microphone output voltages.

The AAA spectrograms of the AH recordings at the eardrum and at the entrance to the blocked auditory canal show similar global structures (figures 4.1 and 4.2). As expected, the first auditory canal resonance in the frequency range between 3 and 5 kHz is prominent in the eardrum recording (figure 4.2), but reduced in the blocked auditory canal situation (figure 4.1). Both systems are according to definition 4 spectro-temporally effective.

4.4 Headphone Playback Situation

The signals represented by the sequences $\mathbf{s}_{\text{hp}}$ are in a typical BS playback situation (subscript *play*) presented using HPs (subscript *hp*). Thereby, the playback path transfer characteristics are superimposed on the signals played back. Instead of HPs, two or more LSs can be used for the playback if the synthesized ear signals are ensured not to reach the respective contra-lateral ear (Møller 1988, 1989). This so-called crosstalk cancellation procedure is discussed in section 3.1.2, while the HP presentation is addressed here.

The HP playback paths include, in addition to the output equipment according to equation 2.8, the paths from the HP input voltages to the ear signals under the HPs (superscript *h*), represented by the TFs

$$\mathbf{H}^{\text{ind,h}}_{\mathbf{u}_{\text{hp}},\mathbf{p}_{\text{e}}}(\mathbf{x}_{\text{hp}_{\text{play}}}) = \frac{\mathbf{P}^{\text{ind,h}}_{\text{e}}(\mathbf{x}_{\text{hp}_{\text{play}}})}{\mathbf{U}_{\text{hp}}}. \tag{4.20}$$

Comparable to the reference scene description, given by equation 4.3 based on the TFs relating the spectrum of the LS input sequence to the ear signal spectra, the playback situation is described, relating the spectra of the HP input sequences to the ear signal spectra under the HPs, by the playback situation TFs

$$\mathbf{H}^{\text{ind,h}}_{\text{play}}(\mathbf{x}_{\text{hp}_{\text{play}}}) = \frac{\mathbf{P}^{\text{ind,h}}_{\text{e}}(\mathbf{x}_{\text{hp}_{\text{play}}})}{\mathbf{S}_{\text{hp}}} = \mathbf{H}_{\text{o}_{\text{hp}}} \cdot \mathbf{H}^{\text{ind,h}}_{\mathbf{u}_{\text{hp}},\mathbf{p}_{\text{e}}}(\mathbf{x}_{\text{hp}_{\text{play}}}). \tag{4.21}$$

Using one of the recording setups described in section 4.3, approximating the playback situation TFs is possible. Relating the spectra of the corresponding microphone input and HP output sequences allows, by including the microphone and input system characteristics in equation 4.21, for formulating the TFs

$$\mathbf{H}^{\mathrm{ind,h}}_{\mathrm{hptf}}(\mathbf{x}_{\mathrm{hp}}, \mathbf{x}_{\mathrm{mic}}) = \frac{\mathbf{S}^{\mathrm{ind,h}}_{\mathrm{mic}}(\mathbf{x}_{\mathrm{hp}}, \mathbf{x}_{\mathrm{mic}})}{\mathbf{S}_{\mathrm{hp}}}. \tag{4.22}$$

The TFs defined by equation 4.22 are, following definition 26, referred to as headphone transfer functions (HPTFs) with corresponding headphone impulse responses (HPIRs).

Definition 26 (*Headphone Impulse Responses and Transfer Functions*)

The transfer paths from headphone driving sequences to sequences stemming from microphones in or attached to the auditory canals under the headphones are modeled by headphone impulse responses with corresponding headphone transfer functions.

In the following paragraphs, approximating the playback situation TFs is discussed for probe microphone measurement with the probe tube tips in the auditory canals close to the eardrums, miniature microphone measurement at the entrances to the blocked auditory canals, and conventional AH measurement with the microphones at the eardrum positions. Each approximation method is described by the corresponding HPTFs and by relating the approximation results to the actual playback situation TFs.

Implementation problems of the different playback situation approximations may be hardware specific, as for example probe tubes may influence the TFs of different HP models to different degrees, or blocked auditory canal entrance measurements may not be suitable for in-ear HPs worn in the canal. However, since this section aims at a hardware independent theoretical discussion, all situations are formulated independent of the HPs.

Probe Microphone Playback Situation Approximation The probe microphone playback situation approximation is described using the microphone TFs defined by equation 2.6 by

$$\begin{aligned}\mathbf{H}^{\mathrm{ind,h}}_{\mathrm{play}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{play}}}) &\approx \mathbf{H}^{\mathrm{ind,h}}_{\mathrm{play}_{\mathrm{pm}}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{play}}}, \mathbf{x}_{\mathrm{pm}_{\mathrm{play}}}) = \mathbf{H}_{\mathrm{o}_{\mathrm{hp}}} \cdot \mathbf{H}^{\mathrm{ind,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{pm}}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{play}}}, \mathbf{x}_{\mathrm{pm}_{\mathrm{play}}}) \\ &= \frac{\mathbf{P}^{\mathrm{ind,h}}_{\mathrm{pm}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{play}}}, \mathbf{x}_{\mathrm{pm}_{\mathrm{play}}})}{\mathbf{S}_{\mathrm{hp}}} = \frac{\mathbf{S}^{\mathrm{ind,h}}_{\mathrm{pm}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{play}}}, \mathbf{x}_{\mathrm{pm}_{\mathrm{play}}})}{\mathbf{S}_{\mathrm{hp}} \cdot \mathbf{H}_{\mathrm{pm}} \cdot \mathbf{H}_{\mathrm{i}_{\mathrm{pm}}}}.\end{aligned} \tag{4.23}$$

The probe microphone HPTFs are given combining equations 4.22 and 4.23 by

$$\mathbf{H}^{\mathrm{ind,h}}_{\mathrm{hptf}_{\mathrm{pm}}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{hptf}}}, \mathbf{x}_{\mathrm{pm}_{\mathrm{hptf}}}) = \mathbf{H}_{\mathrm{o}_{\mathrm{hp}}} \cdot \mathbf{H}^{\mathrm{ind,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{pm}}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{hptf}}}, \mathbf{x}_{\mathrm{pm}_{\mathrm{hptf}}}) \cdot \mathbf{H}_{\mathrm{pm}} \cdot \mathbf{H}_{\mathrm{i}_{\mathrm{pm}}}. \tag{4.24}$$

Blocked Auditory Canal Playback Situation Approximation Blocked auditory canal HP measurements are virtually impossible with consumer in-ear HPs. Therefore, this paragraph primarily applies to extra-aural HPs. If the playback situation is described by miniature microphone measurements at the entrances to the blocked auditory canals, the sound pressure signals at the microphones deviate from the ear signals. Furthermore, the ear signals are not present when the auditory canals are blocked. Therefore, it is impossible to directly measure the so-called blocking factors for HP reproduction

$$\mathbf{H}^{\mathrm{ind,h}}_{\mathbf{p}_{\mathrm{m_b}},\mathbf{p}_{\mathrm{e}},\mathrm{hp}}(\mathbf{x}_{\mathrm{hp}},\mathbf{x}_{\mathrm{hp_b}},\mathbf{x}_{\mathrm{m_b}}) = \frac{\mathbf{P}^{\mathrm{ind,h}}_{\mathrm{e}}(\mathbf{x}_{\mathrm{hp}})}{\mathbf{P}^{\mathrm{ind,b,h}}_{\mathrm{m}}(\mathbf{x}_{\mathrm{hp_b}},\mathbf{x}_{\mathrm{m_b}})} = \frac{\mathbf{H}^{\mathrm{ind,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{e}}}(\mathbf{x}_{\mathrm{hp}})}{\mathbf{H}^{\mathrm{ind,h,b}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{m}}}(\mathbf{x}_{\mathrm{hp_b}},\mathbf{x}_{\mathrm{m_b}})}, \tag{4.25}$$

which relate the ear signal spectra to the sound pressure spectra detected by the miniature microphones at the blocked auditory canal entrances. However, combining the microphone TFs described by equation 2.6 with the playback situation TFs given by equation 4.21 and the blocking factors defined by equation 4.25 allows for the formulation of the TFs

$$\begin{aligned}\mathbf{H}^{\mathrm{ind,h}}_{\mathrm{play}}(\mathbf{x}_{\mathrm{hp_{play}}}) &= \\ &= \frac{\mathbf{S}^{\mathrm{ind,b,h}}_{\mathrm{m}}(\mathbf{x}_{\mathrm{hp_{play,b}}},\mathbf{x}_{\mathrm{m_{play,b}}})}{\mathbf{S}_{\mathrm{hp}}\cdot\mathbf{H}_{\mathrm{m}}\cdot\mathbf{H}_{\mathrm{i_m}}}\cdot\mathbf{H}^{\mathrm{ind,h}}_{\mathbf{p}_{\mathrm{m_b}},\mathbf{p}_{\mathrm{e}},\mathrm{hp}}(\mathbf{x}_{\mathrm{hp_{play}}},\mathbf{x}_{\mathrm{hp_{play,b}}},\mathbf{x}_{\mathrm{m_{play,b}}}),\end{aligned} \tag{4.26}$$

describing the playback situation approximation by blocked auditory canal entrance miniature microphone measurement. The corresponding blocked auditory canal HPTFs are formulated using equations 4.21, 4.22, and 4.26 by

$$\mathbf{H}^{\mathrm{ind,h,b}}_{\mathrm{hptf_m}}(\mathbf{x}_{\mathrm{hp_{hptf}}},\mathbf{x}_{\mathrm{m_{hptf}}}) = \mathbf{H}_{\mathrm{o_{hp}}}\cdot\frac{\mathbf{H}^{\mathrm{ind,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{e}}}(\mathbf{x}_{\mathrm{hp_{play}}})}{\mathbf{H}^{\mathrm{ind,h}}_{\mathbf{p}_{\mathrm{m_b}},\mathbf{p}_{\mathrm{e}},\mathrm{hp}}(\mathbf{x}_{\mathrm{hp_{play}}},\mathbf{x}_{\mathrm{hp_{hptf}}},\mathbf{x}_{\mathrm{m_{hptf}}})}\cdot\mathbf{H}_{\mathrm{m}}\cdot\mathbf{H}_{\mathrm{i_m}}. \tag{4.27}$$

Figure 4.3 shows the AAA spectrogram of a blocked auditory canal AH HPTF[2] according to equation 4.27. In this case, the AH may be regarded as a specific human head.

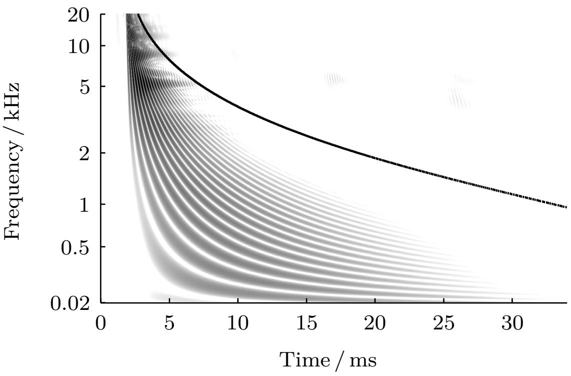

Figure 4.3: Auditory-adapted analysis spectrogram with 60 dB visible dynamic of the transfer path between a headphone and a microphone at the entrance to the blocked auditory canal of an artificial head. The transfer path is defined between the digital sample sequences corresponding to the headphone input and to the microphone output voltages. The black contour indicates the monaural temporal resolution (cf. section 2.4).

Similar to the AAA spectrogram of the blocked auditory canal AH recording situation TF with LS reproduction depicted by figure 4.1, the AAA spectrogram of the blocked auditory

[2] Sennheiser HD 800 headphone, custom-made artificial head $\mathrm{AH_c}$

canal AH HPTF shows a rather frequency independent spectral region ranging from the lower limit of the transmission bandwidth to about 3 kHz. Below that frequency, the HP shows a less frequency dependent phase than the LS. At higher frequencies, spurious maxima occur, according to Pralong and Carlile (1996) due to resonances of the HP-ear system. The blocked auditory canal HPTF shown by figure 4.3 decays by 60 dB within the monaural temporal resolution, indicated by the black contour, at frequencies below about 5 kHz, and is therefore considered purely spectrally effective. For that reason, the HPTF is fully described by the transfer characteristics depicted by figure 4.4.

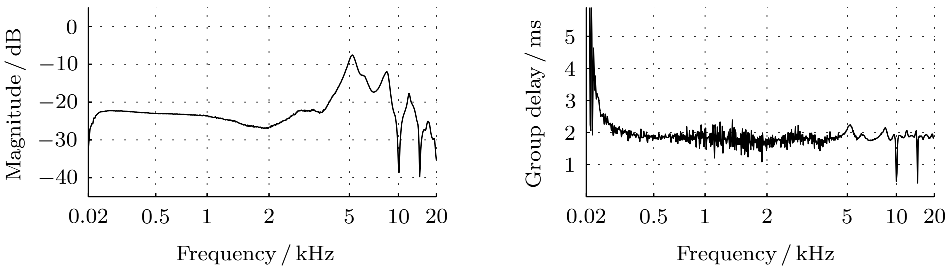

Figure 4.4: Transfer characteristics of the transfer path between a headphone and a microphone at the entrance to the blocked auditory canal of an artificial head. The transfer path is defined between the digital sample sequences corresponding to the headphone input and to the microphone output voltages.

The measurement results depicted by figure 4.4 confirm the conclusions derived based on the AAA spectrogram shown by figure 4.3. In the present case, the temporal system characteristics additionally revealed by the AAA spectrogram are according to section 2.4 inaudible and therefore disrupt an auditory-adapted presentation. The transfer characteristics, on the contrary, aim at representing only the audible impact of the spectrally effective system, and therefore visualize the relevant details more clearly and similarly to the frequently used magnitude and group delay display of standard system TFs.

Artificial Head Playback Situation Approximation The AH playback situation approximation is formulated using the microphone TFs given by equation 2.6 and the playback situation TFs defined by equation 4.21 by the TFs

$$\mathbf{H}^{\mathrm{ind,h}}_{\mathrm{play}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{play}}}) \approx \mathbf{H}^{\mathrm{ah,h}}_{\mathrm{play}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{play}}}) = \frac{\mathbf{S}^{\mathrm{ah,h}}_{\mathrm{ahm}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{play}}})}{\mathbf{S}_{\mathrm{hp}} \cdot \mathbf{H}_{\mathrm{ahm}} \cdot \mathbf{H}_{\mathrm{i}_{\mathrm{ah}}}} = \mathbf{H}_{\mathrm{o}_{\mathrm{hp}}}\,\mathbf{H}^{\mathrm{ah,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{play}}}). \tag{4.28}$$

The corresponding HPTFs are computed combining equations 4.22 and 4.28 to

$$\mathbf{H}^{\mathrm{ah,h}}_{\mathrm{hptf}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{hptf}}}) = \mathbf{H}_{\mathrm{o}_{\mathrm{hp}}} \cdot \mathbf{H}^{\mathrm{ah,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{hptf}}}) \cdot \mathbf{H}_{\mathrm{ahm}} \cdot \mathbf{H}_{\mathrm{i}_{\mathrm{ah}}}. \tag{4.29}$$

Figure 4.5 shows the transfer characteristics representing an AH HPTF with microphones at the eardrum positions measured in the situation also shown by figure 4.4.

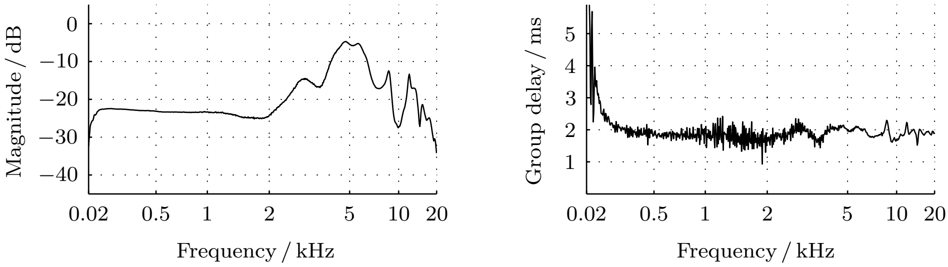

Figure 4.5: Transfer characteristics of the path between a headphone and an artificial head microphone in the situation and using the hardware of figure 4.4, but with the microphone located at the eardrum position.

The results of the eardrum and blocked auditory canal measurements (figures 4.4 and 4.5) differ primarily in the more pronounced first auditory canal resonance of the eardrum measurement visible in figure 4.5 in the frequency range around 5 kHz. This frequency is somewhat higher than expected for human subjects (cf. Fastl and Zwicker 2007, p. 21), due to the compared to average human dimensions smaller AH auditory canal. Figure 4.6 shows the AAA spectrogram corresponding to figure 4.5.

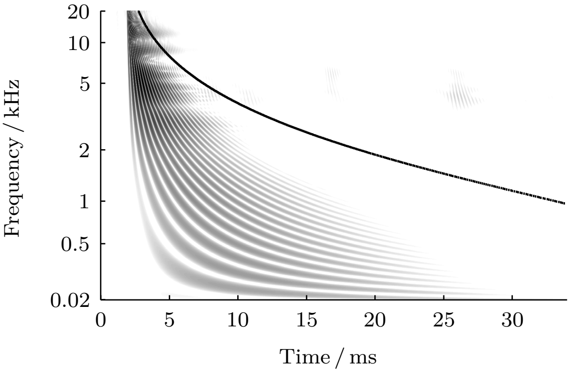

Figure 4.6: Auditory-adapted analysis spectrogram with 60 dB visible dynamic of the transfer path between a headphone and an artificial head microphone in the situation and using the hardware of figure 4.3, but with the microphone located at the eardrum position. The black contour indicates the monaural temporal resolution.

Discussing the HPTF based on the transfer characteristics is justified by the AAA spectrogram depicted by figure 4.6. A comparison to the monaural temporal resolution indicates primarily spectral effectiveness, especially at frequencies below about 7 kHz.

4.5 Non-Equalized Binaural Synthesis

Convolving an audio signal with the BIRPs recorded according to section 4.3 and presenting the convolution products using HPs as described in section 4.4 is referred to as the (temporary) non-equalized BS situation (subscript *ne*). Combining the recording situation TFs given by equation 4.12, recorded according to assumption 1 in the reference scene, with the playback situation TFs given by equation 4.21 allows formulating the non-equalized BS situation ear signals spectra

$$\begin{aligned}\mathbf{P}_{\mathrm{e_{ne}}}^{\mathrm{ind,h}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{mic_{rec}}}, \mathbf{x}_{\mathrm{hp_{play}}}) = \\ = S_{\mathrm{ls}} \cdot \mathbf{H}_{\mathrm{rec_{mic}}}^{\mathrm{ind}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{mic_{rec}}}) \cdot \mathbf{H}_{\mathrm{play}}^{\mathrm{ind,h}}(\mathbf{x}_{\mathrm{hp_{play}}}).\end{aligned} \tag{4.30}$$

Comparing equation 4.30 to the reference scene ear signal spectra given by equation 4.2 reveals different ear signals in the non-equalized BS situation and the reference scene. However, all combinations of the recording and playback situations are discussed, serving as the basis for the derivation of the associated equalization requirements in section 4.6. For every combination of the recording and playback situations, the TFs

$$\begin{aligned}\mathbf{H}_{\mathrm{ne}}^{\mathrm{ind,h}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{mic_{rec}}}, \mathbf{x}_{\mathrm{hp_{play}}}) &= \frac{\mathbf{P}_{\mathrm{e_{ne}}}^{\mathrm{ind,h}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{mic_{rec}}}, \mathbf{x}_{\mathrm{hp_{play}}})}{S_{\mathrm{ls}}} \\ &= \mathbf{H}_{\mathrm{rec_{mic}}}^{\mathrm{ind}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{mic_{rec}}}) \cdot \mathbf{H}_{\mathrm{play}}^{\mathrm{ind,h}}(\mathbf{x}_{\mathrm{hp_{play}}})\end{aligned} \tag{4.31}$$

are formulated, connecting the source sequence and ear signal spectra. Probe microphone, blocked auditory canal miniature microphone, and artificial head recording are discussed in the following for human and artificial head playback.

4.5.1 Non-Equalized Human Head Playback

Since the goal of BS is the acoustic simulation of a reference scene, the synthesized ear signals are typically presented to a human listener. Therefore, equation 4.31 is adapted to human head playback for each recording method in this section.

Human Head Playback and Probe Microphone Recording According to assumption 1, the reference scene equipment is used for the recording. Consequently, by inserting the probe microphone recording situation TFs of equation 4.13 and the human head playback situation TFs given by equation 4.21, equation 4.31 can be specified to the TFs

$$\begin{aligned}&\mathbf{H}_{\mathrm{ne_{pm}}}^{\mathrm{ind,h}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{pm_{rec}}}, \mathbf{x}_{\mathrm{hp_{play}}}) = \\ &= H_{\mathrm{o_{ref}}} \cdot \mathbf{H}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{pm}}}^{\mathrm{ind}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{pm_{rec}}}) \cdot \mathbf{H}_{\mathrm{pm}} \cdot \mathbf{H}_{\mathrm{i_{pm}}} \cdot \mathbf{H}_{\mathrm{o_{hp}}} \cdot \mathbf{H}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{e}}}^{\mathrm{ind,h}}(\mathbf{x}_{\mathrm{hp_{play}}}),\end{aligned} \tag{4.32}$$

connecting the source signal and ear signal spectra for probe microphone recording. These TFs are formulated by the following equation with respect to the reference scene, pointing out deviations from the intended reference scene reproduction. Assuming the sound pressure signals at the probe microphones represent the ear signals, the combination of equation 4.32 with the relation of recording situation and reference scene given by equation 4.14 results in

$$\begin{aligned}\mathbf{H}_{\mathrm{ne_{pm}}}^{\mathrm{ind,h}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{pm_{rec}}}, \mathbf{x}_{\mathrm{hp_{play}}}) \approx \\ \approx \mathbf{H}_{\mathrm{ref}}^{\mathrm{ind}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}) \cdot \mathbf{H}_{\mathrm{pm}} \cdot \mathbf{H}_{\mathrm{i_{pm}}} \cdot \mathbf{H}_{\mathrm{o_{hp}}} \cdot \mathbf{H}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{e}}}^{\mathrm{ind,h}}(\mathbf{x}_{\mathrm{hp_{play}}}).\end{aligned} \tag{4.33}$$

Human Head Playback and Blocked Auditory Canal Recording The TFs connecting source and ear signal spectra for blocked auditory canal recording and human head playback are derived comparably. Combining equations 4.16, 4.21, and 4.31 results in

$$\begin{aligned}\mathbf{H}^{\text{ind,h}}_{\text{ne}_\text{m}}(\mathbf{x}_{\text{h}_\text{ref}},\mathbf{x}_{\text{ls}_\text{ref}},\mathbf{x}_{\text{m}_\text{rec}},\mathbf{x}_{\text{hp}_\text{play}}) &= \mathbf{H}^{\text{ind}}_{\text{ref}}(\mathbf{x}_{\text{h}_\text{ref}},\mathbf{x}_{\text{ls}_\text{ref}})\cdot \\ &\cdot \frac{\mathbf{H}^{\text{ind,b}}_{u_\text{ls},\mathbf{p}_\text{m}}(\mathbf{x}_{\text{h}_\text{ref}},\mathbf{x}_{\text{ls}_\text{ref}},\mathbf{x}_{\text{m}_\text{rec}})}{\mathbf{H}^{\text{ind}}_{u_\text{ls},\mathbf{p}_\text{e}}(\mathbf{x}_{\text{h}_\text{ref}},\mathbf{x}_{\text{ls}_\text{ref}})}\cdot\mathbf{H}_\text{m}\cdot\mathbf{H}_{\text{i}_\text{m}}\cdot\mathbf{H}_{\text{o}_\text{hp}}\cdot\mathbf{H}^{\text{ind,h}}_{\mathbf{u}_\text{hp},\mathbf{p}_\text{e}}(\mathbf{x}_{\text{hp}_\text{play}}).\end{aligned} \tag{4.34}$$

Human Head Playback and Artificial Head Recording The derivation given for probe microphone recording and human head playback holds also true for artificial head recording and human head playback. The result is given with equations 4.19, 4.21, and 4.31 by

$$\begin{aligned}\mathbf{H}^{\text{ind,h}}_{\text{ne}_\text{ah}}(\mathbf{x}_{\text{h}_\text{ref}},\mathbf{x}_{\text{ls}_\text{ref}},\mathbf{x}_{\text{hp}_\text{play}}) &= \mathbf{H}^{\text{ind}}_{\text{ref}}(\mathbf{x}_{\text{h}_\text{ref}},\mathbf{x}_{\text{ls}_\text{ref}})\cdot \\ &\cdot \frac{\mathbf{H}^{\text{ah}}_{u_\text{ls},\mathbf{p}_\text{ahm}}(\mathbf{x}_{\text{h}_\text{ref}},\mathbf{x}_{\text{ls}_\text{ref}})}{\mathbf{H}^{\text{ind}}_{u_\text{ls},\mathbf{p}_\text{e}}(\mathbf{x}_{\text{h}_\text{ref}},\mathbf{x}_{\text{ls}_\text{ref}})}\cdot\mathbf{H}_\text{ahm}\cdot\mathbf{H}_{\text{i}_\text{ah}}\cdot\mathbf{H}_{\text{o}_\text{hp}}\cdot\mathbf{H}^{\text{ind,h}}_{\mathbf{u}_\text{hp},\mathbf{p}_\text{e}}(\mathbf{x}_{\text{hp}_\text{play}}).\end{aligned} \tag{4.35}$$

Before the equalization requirements for BS are derived in section 4.6 based on the TFs of the non-equalized scenarios given in the present section, the AH playback situations are formulated. In order to allow for a comparison of the recording procedures, each of the three recording methods defined in section 4.3 is taken into account.

4.5.2 Non-Equalized Artificial Head Playback

AH playback is, due to its simplicity and reproducibility, a helpful and frequently used step in the design and implementation process of BS. For that reason, AH playback is discussed in this section, assuming the AH reference scene approximation given by equation 4.11 as the reference scene.

Artificial Head Playback and Probe Microphone Recording For probe microphone recording, the TFs relating the source and ear signal spectra can be formulated based on the non-equalized BS situation given by equation 4.31. Using the recording situation TFs defined by 4.13 and the AH playback situation TFs described by equation 4.28 reveals

$$\begin{aligned}&\mathbf{H}^{\text{ah,h}}_{\text{ne}_\text{pm}}(\mathbf{x}_{\text{h}_\text{ref}},\mathbf{x}_{\text{ls}_\text{ref}},\mathbf{x}_{\text{pm}_\text{rec}},\mathbf{x}_{\text{hp}_\text{play}}) = \\ &= H_{\text{o}_\text{ref}}\cdot\mathbf{H}^{\text{ind}}_{u_\text{ls},\mathbf{p}_\text{pm}}(\mathbf{x}_{\text{h}_\text{ref}},\mathbf{x}_{\text{ls}_\text{ref}},\mathbf{x}_{\text{pm}_\text{rec}})\cdot\mathbf{H}_\text{pm}\cdot\mathbf{H}_{\text{i}_\text{pm}}\cdot\mathbf{H}_{\text{o}_\text{hp}}\cdot\mathbf{H}^{\text{ah,h}}_{\mathbf{u}_\text{hp},\mathbf{p}_\text{ahm}}(\mathbf{x}_{\text{hp}_\text{play}}).\end{aligned} \tag{4.36}$$

Equation 4.36 can be given in relation to the AH reference scene by comparison to the reference scene descriptions in equations 4.10 and 4.11. This procedure results in

$$\begin{aligned}\mathbf{H}^{\text{ah,h}}_{\text{ne}_\text{pm}}(\mathbf{x}_{\text{h}_\text{ref}},\mathbf{x}_{\text{ls}_\text{ref}},\mathbf{x}_{\text{pm}_\text{rec}},\mathbf{x}_{\text{hp}_\text{play}}) &= \mathbf{H}^{\text{ah}}_{\text{ref}_\text{ahm}}(\mathbf{x}_{\text{h}_\text{ref}},\mathbf{x}_{\text{ls}_\text{ref}})\cdot \\ &\cdot \frac{\mathbf{H}^{\text{ind}}_{u_\text{ls},\mathbf{p}_\text{pm}}(\mathbf{x}_{\text{h}_\text{ref}},\mathbf{x}_{\text{ls}_\text{ref}},\mathbf{x}_{\text{pm}_\text{rec}})}{\mathbf{H}^{\text{ah}}_{u_\text{ls},\mathbf{p}_\text{ahm}}(\mathbf{x}_{\text{h}_\text{ref}},\mathbf{x}_{\text{ls}_\text{ref}})}\cdot\mathbf{H}_\text{pm}\cdot\mathbf{H}_{\text{i}_\text{pm}}\cdot\mathbf{H}_{\text{o}_\text{hp}}\cdot\mathbf{H}^{\text{ah,h}}_{\mathbf{u}_\text{hp},\mathbf{p}_\text{ahm}}(\mathbf{x}_{\text{hp}_\text{play}}).\end{aligned} \tag{4.37}$$

Artificial Head Playback and Blocked Auditory Canal Recording The combination of blocked auditory canal recording and AH playback is described comparable to the probe microphone situation. Using equations 4.15, 4.28, and 4.31 allows formulating

$$\begin{aligned}&\mathbf{H}^{\mathrm{ah,h}}_{\mathrm{ne_m}}(\mathbf{x}_{\mathrm{h_{ref}}},\mathbf{x}_{\mathrm{ls_{ref}}},\mathbf{x}_{\mathrm{m_{rec}}},\mathbf{x}_{\mathrm{hp_{play}}}) = \\ &\quad = H_{\mathrm{o_{ref}}} \cdot \mathbf{H}^{\mathrm{ind,b}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{m}}}(\mathbf{x}_{\mathrm{h_{ref}}},\mathbf{x}_{\mathrm{ls_{ref}}},\mathbf{x}_{\mathrm{m_{rec}}}) \cdot \mathbf{H}_{\mathrm{m}} \cdot \mathbf{H}_{\mathrm{i_m}} \cdot \mathbf{H}_{\mathrm{o_{hp}}} \cdot \mathbf{H}^{\mathrm{ah,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{hp_{play}}}).\end{aligned} \tag{4.38}$$

The relation of equation 4.38 to the AH reference scene is given by comparison to the reference scene descriptions in equations 4.10 and 4.11. This procedure results in

$$\begin{aligned}&\mathbf{H}^{\mathrm{ah,h}}_{\mathrm{ne_m}}(\mathbf{x}_{\mathrm{h_{ref}}},\mathbf{x}_{\mathrm{ls_{ref}}},\mathbf{x}_{\mathrm{m_{rec}}},\mathbf{x}_{\mathrm{hp_{play}}}) = \mathbf{H}^{\mathrm{ah}}_{\mathrm{ref_{ahm}}}(\mathbf{x}_{\mathrm{h_{ref}}},\mathbf{x}_{\mathrm{ls_{ref}}}) \cdot \\ &\quad \cdot \frac{\mathbf{H}^{\mathrm{ind,b}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{m}}}(\mathbf{x}_{\mathrm{h_{ref}}},\mathbf{x}_{\mathrm{ls_{ref}}},\mathbf{x}_{\mathrm{m_{rec}}})}{\mathbf{H}^{\mathrm{ah}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{h_{ref}}},\mathbf{x}_{\mathrm{ls_{ref}}})} \cdot \mathbf{H}_{\mathrm{m}} \cdot \mathbf{H}_{\mathrm{i_m}} \cdot \mathbf{H}_{\mathrm{o_{hp}}} \cdot \mathbf{H}^{\mathrm{ah,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{hp_{play}}}).\end{aligned} \tag{4.39}$$

Artificial Head Playback and Artificial Head Recording The most simple BS evaluation procedure employs an AH for recording and playback. Inserting equations 4.18 and 4.28 in equation 4.31, the AH presentation of AH recordings is described by

$$\begin{aligned}&\mathbf{H}^{\mathrm{ah,h}}_{\mathrm{ne_{ah}}}(\mathbf{x}_{\mathrm{h_{ref}}},\mathbf{x}_{\mathrm{ls_{ref}}},\mathbf{x}_{\mathrm{hp_{play}}}) = \\ &\quad = \mathbf{H}^{\mathrm{ah}}_{\mathrm{ref_{ahm}}}(\mathbf{x}_{\mathrm{h_{ref}}},\mathbf{x}_{\mathrm{ls_{ref}}}) \cdot \mathbf{H}_{\mathrm{ahm}} \cdot \mathbf{H}_{\mathrm{i_{ah}}} \cdot \mathbf{H}_{\mathrm{o_{hp}}} \cdot \mathbf{H}^{\mathrm{ah,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{hp_{play}}}).\end{aligned} \tag{4.40}$$

Based on the formalism introduced in this section, equalization requirements for system-theoretically correct BS are formulated in the next section. The way of presentation is selected with the aim of reflecting the sources and amounts of errors in the resulting mathematical description. Equalization requirements are formulated for all combinations of recording and playback methods discussed, to provide a systematic and theoretically well-defined comparison of the synthesis results achievable with the different procedures.

4.6 Equalization Requirements for Binaural Synthesis

Reproducing the reference scene ear signals, which are by definition 22 the sound pressure signals detected by the eardrums, is frequently assumed sufficient for recreating the reference scene hearing sensations (Møller 1992). Neglecting non-acoustic effects (cf. section 3.2.1), this requirement seems satisfactory since the human ear acts according to Fastl and Zwicker (2007, p. 25) in good approximation as a pressure receiver. However, some authors report different auditory canal levels at equal loudness for HP versus LS reproduction (e. g. Beranek 1949, Fastl et al. 1985). Stimulating only the auditory modality synthetically while keeping the overall listening situation constant, the validity of Møller's BS design goal is addressed by loudness comparisons in section 5.4. Based on the results, the recreation of the reference scene ear signals is assumed as a sufficient goal for BS (assumption 2, cf. Völk et al. 2011d, Völk and Fastl 2011a).

Assumption 2 (*Ear Signals as the Binaural Synthesis Design Goal*)

Keeping the overall situation apart from the sound sources constant, the recreation of the reference scene ear signals is sufficient to recreate the corresponding auditory impressions and is therefore the design goal for binaural synthesis.

Mathematically, the BS design goal of recreating the reference scene ear signals is imposed on the binaurally synthesized ear signal spectra by

$$\mathbf{P}_{\mathrm{e_{bs}}}^{\mathrm{ind,h}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{mic_{rec}}}, \mathbf{x}_{\mathrm{hp_{play}}}) \overset{!}{=} \mathbf{P}_{\mathrm{e}}^{\mathrm{ind}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}). \tag{4.41}$$

In order to reach this goal, according to section 4.5 and taking into account equations 4.2, 4.3, and 4.30, equalization filters must be applied so that

$$\mathbf{H}_{\mathrm{ne}}^{\mathrm{ind,h}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{mic_{rec}}}, \mathbf{x}_{\mathrm{hp_{play}}}) \cdot \mathbf{H}_{\mathrm{eq}}^{\mathrm{ind}}(\mathbf{x}_{\mathrm{hp_{play}}}, \mathbf{x}_{\mathrm{mic_{rec}}}) = \mathbf{H}_{\mathrm{ref}}^{\mathrm{ind}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}) \tag{4.42}$$

holds true. Assuming the invertibility of the non-equalized BS TFs (assumption 3), the equalization filters are in general computed by

$$\mathbf{H}_{\mathrm{eq}}^{\mathrm{ind}}(\mathbf{x}_{\mathrm{hp_{play}}}, \mathbf{x}_{\mathrm{mic_{rec}}}) = \frac{\mathbf{H}_{\mathrm{ref}}^{\mathrm{ind}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}})}{\mathbf{H}_{\mathrm{ne}}^{\mathrm{ind,h}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{mic_{rec}}}, \mathbf{x}_{\mathrm{hp_{play}}})}. \tag{4.43}$$

Assumption 3 (*Invertibility of Non-Equalized Transfer Functions*)

The transfer functions describing the non-equalized binaural synthesis are invertible.

In the following, the equalization filters for system-theoretically correct BS are specified to the situations defined in section 4.5. The results are related to the HPTFs introduced in section 4.4 since BS equalization filters are commonly designed based on inverted HPTFs.

4.6.1 Equalization Requirements for Human Head Playback

According to definition 1, formulating the aim of virtual acoustics procedures in general, the purpose of BS as a virtual acoustics procedure is to elicit the reference scene hearing sensations. Therefore, the playback to individual subjects is discussed in this section, taking into account all situations introduced in section 4.5. After specifying the equalization requirements for each combination of playback and recording situation, the equalization results achievable using the HPTFs introduced in section 4.4 are formulated.

Human Head Playback and Probe Microphone Recording

If the equalization of probe microphone recordings with the probe tube tips close to the eardrums is attempted, it is necessary to consider the sound pressure signals at the probe tube tips only approximate the ear signals (section 5.1.4). However, the theory is derived at first assuming the probe tube tip pressures represent the ear signals (assumption 4).

Assumption 4 (*Probe Microphone Sound Pressure and Ear Signals*)

> *The sound pressure signals detected by probe microphones with the probe tube tips close to the eardrums represent the ear signals.*

The equalization requirements given by equation 4.43 can be specified for the human head playback of probe microphone recordings using equation 4.33. This procedure results in

$$\mathbf{H}^{\text{ind}}_{\text{eq}_{\text{pm}}}(\mathbf{x}_{\text{hp}_{\text{play}}}, \mathbf{x}_{\text{pm}_{\text{rec}}}) \approx \frac{1}{\mathbf{H}_{\text{pm}} \cdot \mathbf{H}_{\text{i}_{\text{pm}}} \cdot \mathbf{H}_{\text{o}_{\text{hp}}} \cdot \mathbf{H}^{\text{ind,h}}_{\mathbf{u}_{\text{hp}},\mathbf{p}_{\text{e}}}(\mathbf{x}_{\text{hp}_{\text{play}}})}. \tag{4.44}$$

In the following, the requirements defined by equation 4.44 are given in relation to the HPTF measurement procedures defined in section 4.4.

Human Head Playback, Probe Microphone Recording and Headphone Transfer Functions Combined with equation 4.24 and assumption 4, equation 4.44 results in a description of the equalization requirements for probe microphone measurement and human head playback dependent on probe microphone HPTFs. This description is given by

$$\mathbf{H}^{\text{ind}}_{\text{eq}_{\text{pm}}}(\mathbf{x}_{\text{hp}_{\text{play}}}, \mathbf{x}_{\text{pm}_{\text{rec}}}) \approx \frac{1}{\mathbf{H}^{\text{ind,h}}_{\text{hptf}_{\text{pm}}}(\mathbf{x}_{\text{hp}_{\text{hptf}}}, \mathbf{x}_{\text{pm}_{\text{hptf}}})}. \tag{4.45}$$

Equation 4.45 reveals the equalization requirements for system-theoretically correct BS if probe microphones with the tube tips close to the eardrums are used for all measurements. Consequently, equalizing a BS system with inverted HPTFs is possible under certain conditions. It must be ensured that the probe tube tips are positioned identically during HPTF measurement and recording (assumption 5).

Assumption 5 (*Microphone Positions*)

> *The microphone positions are identical in the recording and headphone transfer function measurement situations.*

Furthermore, identical HP positions during playback and HPTF measurement would be necessary. This requirement is formulated by assumption 6.

Assumption 6 (*Headphone Positions*)

> *The headphone positions are identical in the playback and headphone transfer function measurement situations.*

Summarizing, if assumptions 4, 5, and 6 are fulfilled, the ear signals generated by BS based on probe microphone measurement equal the reference scene ear signals. While probe microphone measurements provide the most direct way to BS, their results can never be proven by measurement since ear signal measurements in a strict sense are impossible (cf. section 5.1.4). However, equation 4.45 confirms that correctly equalized probe microphone based BS is independent from the microphones and input systems.

Human Head Playback, Probe Microphone Recording, and Blocked Auditory Canal Headphone Transfer Functions In order to describe the relation of the equalization requirements for probe microphone recording and human head playback to blocked auditory canal HPTFs, equation 4.27 and equation 4.44 are combined to

$$\begin{aligned}\mathbf{H}^{\mathrm{ind}}_{\mathrm{eq}_{\mathrm{pm}}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{play}}},\mathbf{x}_{\mathrm{pm}_{\mathrm{rec}}}) \approx{}& \frac{1}{\mathbf{H}^{\mathrm{ind,h,b}}_{\mathrm{hptf}_{\mathrm{m}}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{hptf}}},\mathbf{x}_{\mathrm{m}_{\mathrm{hptf}}})}\cdot\frac{\mathbf{H}_{\mathrm{m}}\cdot\mathbf{H}_{\mathrm{i}_{\mathrm{m}}}}{\mathbf{H}_{\mathrm{pm}}\cdot\mathbf{H}_{\mathrm{i}_{\mathrm{pm}}}}\cdot\\ &\cdot\frac{1}{\mathbf{H}^{\mathrm{ind,h}}_{\mathbf{p}_{\mathrm{m_b}},\mathbf{p}_{\mathrm{e}},\mathrm{hp}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{play}}},\mathbf{x}_{\mathrm{hp}_{\mathrm{hptf}}},\mathbf{x}_{\mathrm{m}_{\mathrm{hptf}}})}.\end{aligned} \tag{4.46}$$

Equation 4.46 reveals that the equalization of BS with probe microphone recording for human head playback is impossible with HPTFs measured at the blocked auditory canal entrances, even if differences in HP and microphone positions are neglected. Errors result from different sound pressures under the HPs at the eardrums and at the miniature microphones at the entrances to the blocked auditory canals. Further, differences of the HP and microphone positions during recording, HPTF measurement, and playback as well as different input systems and microphones can contribute to inaccuracies.

Human Head Playback, Probe Microphone Recording, and Artificial Head Headphone Transfer Functions Comparing equations 4.29 and 4.44, the relation of the equalization requirements for probe microphone recording to AH HPTFs is given to

$$\mathbf{H}^{\mathrm{ind}}_{\mathrm{eq}_{\mathrm{pm}}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{play}}},\mathbf{x}_{\mathrm{pm}_{\mathrm{rec}}}) \approx \frac{1}{\mathbf{H}^{\mathrm{ah,h}}_{\mathrm{hptf}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{hptf}}})}\cdot\frac{\mathbf{H}^{\mathrm{ah,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{hptf}}})}{\mathbf{H}^{\mathrm{ind,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{e}}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{play}}})}\cdot\frac{\mathbf{H}_{\mathrm{ahm}}\cdot\mathbf{H}_{\mathrm{i}_{\mathrm{ah}}}}{\mathbf{H}_{\mathrm{pm}}\cdot\mathbf{H}_{\mathrm{i}_{\mathrm{pm}}}}. \tag{4.47}$$

Equation 4.47 shows that it is not possible to equalize BS with probe microphone recording for human head playback using AH HPTFs. Errors are caused by deviations between the transfer paths from the HPs to the eardrums in the human and artificial head situations. Additionally, distortion is possible due to different microphones and input equipment and due to diverging HP positions in the HPTF measurement and playback situations.

However, if assumption 6 holds true in that identical HP positions are ensured for playback and HPTF measurement, and if high quality microphones and input equipment are used, the error is determined by differences of the transfer paths between the HPs and the eardrums in the artificial and human head cases. Since ear signal measurements on human heads are possible only approximately (cf. section 5.1.4), it may be desirable to support the evaluation by the audible impact of the remaining error determined by listening experiments, for example by loudness adjustments (cf. section 5.4).

Human Head Playback and Blocked Auditory Canal Recording

In this section, the equalization requirements for human head playback are formulated for BS with blocked auditory canal recording. Furthermore, the results are related to the different HPTFs introduced in section 4.4. Using equations 4.34 and 4.43, the equalization requirements for BS with blocked auditory canal recording are given for human head playback, initially HPTF independently, by

$$\begin{aligned}\mathbf{H}^{\mathrm{ind,b}}_{\mathrm{eq_m}}(\mathbf{x}_{\mathrm{hp_{play}}}, \mathbf{x}_{\mathrm{m_{rec}}}) &= \\ &= \frac{\mathbf{H}^{\mathrm{ind}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{e}}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}})}{\mathbf{H}^{\mathrm{ind,b}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{m}}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{m_{rec}}})} \cdot \frac{1}{\mathbf{H}_{\mathrm{m}} \cdot \mathbf{H}_{\mathrm{i_m}} \cdot \mathbf{H}_{\mathrm{o_{hp}}} \cdot \mathbf{H}^{\mathrm{ind,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{e}}}(\mathbf{x}_{\mathrm{hp_{play}}})}.\end{aligned} \tag{4.48}$$

Human Head Playback, Blocked Auditory Canal Recording, and Probe Microphone Headphone Transfer Functions Combining equation 4.48 with equation 4.24, the equalization requirements for BS with blocked auditory canal recording for human head playback can be written dependent on probe microphone HPTFs. In this case, the requirements are formulated by

$$\begin{aligned}\mathbf{H}^{\mathrm{ind,b}}_{\mathrm{eq_m}}(\mathbf{x}_{\mathrm{hp_{play}}}, \mathbf{x}_{\mathrm{m_{rec}}}) = \frac{1}{\mathbf{H}^{\mathrm{ind,h}}_{\mathrm{hptf_{pm}}}(\mathbf{x}_{\mathrm{hp_{hptf}}}, \mathbf{x}_{\mathrm{pm_{hptf}}})} \cdot \frac{\mathbf{H}^{\mathrm{ind,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{pm}}}(\mathbf{x}_{\mathrm{hp_{hptf}}}, \mathbf{x}_{\mathrm{pm_{hptf}}})}{\mathbf{H}^{\mathrm{ind,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{e}}}(\mathbf{x}_{\mathrm{hp_{play}}})} \cdot \\ \cdot \frac{\mathbf{H}^{\mathrm{ind}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{e}}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}})}{\mathbf{H}^{\mathrm{ind,b}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{m}}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{m_{rec}}})} \cdot \frac{\mathbf{H}_{\mathrm{pm}} \cdot \mathbf{H}_{\mathrm{i_{pm}}}}{\mathbf{H}_{\mathrm{m}} \cdot \mathbf{H}_{\mathrm{i_m}}}.\end{aligned} \tag{4.49}$$

If assumptions 4 and 6 hold true, that is if the sound pressures detected by the probe microphone tube tips represent the ear signals, and if the HP positions during HPTF measurement and playback are identical, equation 4.49 simplifies to

$$\begin{aligned}\mathbf{H}^{\mathrm{ind,b}}_{\mathrm{eq_m}}(\mathbf{x}_{\mathrm{hp_{play}}}, \mathbf{x}_{\mathrm{m_{rec}}}) &\approx \\ &\approx \frac{1}{\mathbf{H}^{\mathrm{ind,h}}_{\mathrm{hptf_{pm}}}(\mathbf{x}_{\mathrm{hp_{hptf}}}, \mathbf{x}_{\mathrm{pm_{hptf}}})} \cdot \frac{\mathbf{H}^{\mathrm{ind}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{e}}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}})}{\mathbf{H}^{\mathrm{ind,b}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{m}}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{m_{rec}}})} \cdot \frac{\mathbf{H}_{\mathrm{pm}} \cdot \mathbf{H}_{\mathrm{i_{pm}}}}{\mathbf{H}_{\mathrm{m}} \cdot \mathbf{H}_{\mathrm{i_m}}}.\end{aligned} \tag{4.50}$$

Equation 4.50 reveals that equalizing a BS system implemented with blocked auditory canal entrance recording is impossible using probe microphone HPTFs only. Errors may be caused by differences of the TFs describing the paths from the LS input signal to the blocked auditory canal entrance sound pressure signals and to the eardrum sound pressure signals. Further inaccuracies may result from different microphones and input systems.

Human Head Playback, Blocked Auditory Canal Recording and Headphone Transfer Functions Comparing equations 4.27 and 4.48, it is possible to compute the relation of the equalization requirements for the human head playback in BS with blocked auditory canal recording to blocked auditory canal HPTFs. These requirements are formulated by

$$\begin{aligned}\mathbf{H}^{\mathrm{ind,b}}_{\mathrm{eq_m}}(\mathbf{x}_{\mathrm{hp_{play}}}, \mathbf{x}_{\mathrm{m_{rec}}}) = \frac{1}{\mathbf{H}^{\mathrm{ind,h,b}}_{\mathrm{hptf_m}}(\mathbf{x}_{\mathrm{hp_{hptf}}}, \mathbf{x}_{\mathrm{m_{hptf}}})} \cdot \frac{\mathbf{H}^{\mathrm{ind}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{e}}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}})}{\mathbf{H}^{\mathrm{ind,b}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{m}}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{m_{rec}}})} \cdot \\ \cdot \frac{1}{\mathbf{H}^{\mathrm{ind,h}}_{\mathbf{p}_{\mathrm{m_b}},\mathbf{p}_{\mathrm{e}},\mathrm{hp}}(\mathbf{x}_{\mathrm{hp_{play}}}, \mathbf{x}_{\mathrm{hp_{hptf}}}, \mathbf{x}_{\mathrm{m_{hptf}}})}.\end{aligned} \tag{4.51}$$

The relations of the sound pressure spectra at the blocked auditory canal entrance microphones to those at the eardrums play an important role for the BS with blocked auditory canal recording. These relations, defined for HP reproduction by equation 4.25, are referred to as blocking factors. While with HP reproduction the blocking factors are necessarily defined for subjects wearing active HPs, with an LS as the sound source, the blocking factors can be measured with or without inactive HPs. The blocking factors for LS reproduction without HPs are defined by

$$\mathbf{H}^{\mathrm{ind}}_{\mathbf{p}_{\mathrm{m_b}},\mathbf{p}_{\mathrm{e}},\mathrm{ls}}(\mathbf{x}_{\mathrm{m_{rec}}}) = \frac{\mathbf{P}^{\mathrm{ind}}_{\mathrm{e}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}})}{\mathbf{P}^{\mathrm{ind,b}}_{\mathrm{m}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{m_{rec}}})} = \frac{\mathbf{H}^{\mathrm{ind}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{e}}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}})}{\mathbf{H}^{\mathrm{ind,b}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{m}}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{m_{rec}}})}. \tag{4.52}$$

Using the blocking factors given by equation 4.52, equation 4.51 can be simplified to

$$\begin{aligned} &\mathbf{H}^{\mathrm{ind,b}}_{\mathrm{eq_m}}(\mathbf{x}_{\mathrm{hp_{play}}}, \mathbf{x}_{\mathrm{m_{rec}}}) = \\ &\quad = \frac{1}{\mathbf{H}^{\mathrm{ind,h,b}}_{\mathrm{hptf_m}}(\mathbf{x}_{\mathrm{hp_{hptf}}}, \mathbf{x}_{\mathrm{m_{hptf}}})} \cdot \frac{\mathbf{H}^{\mathrm{ind}}_{\mathbf{p}_{\mathrm{m_b}},\mathbf{p}_{\mathrm{e}},\mathrm{ls}}(\mathbf{x}_{\mathrm{m_{rec}}})}{\mathbf{H}^{\mathrm{ind,h}}_{\mathbf{p}_{\mathrm{m_b}},\mathbf{p}_{\mathrm{e}},\mathrm{hp}}(\mathbf{x}_{\mathrm{hp_{play}}}, \mathbf{x}_{\mathrm{hp_{hptf}}}, \mathbf{x}_{\mathrm{m_{hptf}}})}. \end{aligned} \tag{4.53}$$

The equalization requirements given by equation 4.53 represent a common BS procedure (cf. Møller 1992, Hammershøi and Møller 2002, Völk and Fastl 2011a). The synthesized ear signals equal those of the reference scene if the assumptions 5, 6, and 7 are fulfilled.

Assumption 7 (*Blocked Auditory Canal Recording*)

> *The transfer functions connecting the sound pressure spectra at miniature microphones in the blocked auditory canals and at the eardrums in the open auditory canals are identical for headphone and loudspeaker playback without headphones.*

In other words, the synthesis results are correct if the probe tube tips are positioned identically during HPTF measurement and recording, if identical HP positions during playback and HPTF measurement are ensured, and if equal blocking factors arise for HP and LS reproduction without HPs. Since the blocking factors for HP reproduction depend on the HP model, also the authenticity of the ear signals generated by BS with blocked auditory canal recording equalized with blocked auditory canal HPTFs is influenced by the specific HP model used (cf. sections 5.3 and 5.4).

Human Head Playback, Blocked Auditory Canal Recording, and Artificial Head Headphone Transfer Functions Equations 4.29 and 4.48 allow for the formulation of the equalization requirements for BS with blocked auditory canal recording for human head playback with respect to AH HPTFs. The resulting requirements are given by

$$\begin{aligned} \mathbf{H}^{\mathrm{ind,b}}_{\mathrm{eq_m}}(\mathbf{x}_{\mathrm{hp_{play}}}, \mathbf{x}_{\mathrm{m_{rec}}}) &= \frac{1}{\mathbf{H}^{\mathrm{ah,h}}_{\mathrm{hptf_{ahm}}}(\mathbf{x}_{\mathrm{hp_{hptf}}})} \cdot \frac{\mathbf{H}^{\mathrm{ah,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{hp_{hptf}}})}{\mathbf{H}^{\mathrm{ind,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{e}}}(\mathbf{x}_{\mathrm{hp_{play}}})} \cdot \\ &\quad \cdot \frac{\mathbf{H}^{\mathrm{ind}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{e}}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}})}{\mathbf{H}^{\mathrm{ind,b}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{m}}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{m_{rec}}})} \cdot \frac{\mathbf{H}_{\mathrm{ahm}} \cdot \mathbf{H}_{\mathrm{i_{ah}}}}{\mathbf{H}_{\mathrm{m}} \cdot \mathbf{H}_{\mathrm{i_m}}}. \end{aligned} \tag{4.54}$$

Equation 4.54 shows that the equalization of BS with blocked auditory canal recording for human head playback is not possible using AH HPTFs. Errors may result from deviations between the human and artificial heads, from the differences of the sound pressures at the entrances to the blocked auditory canals and at the eardrums, and from differences of microphones and input equipment.

Human Head Playback and Artificial Head Recording

The equalization requirements for human head playback are formulated for BS with AH recording in the following. In addition, the results are related to the HPTFs introduced in section 4.4. Combining equations 4.35 and 4.43, the equalization requirements for the human head playback of AH recordings are given in general by

$$\mathbf{H}^{\mathrm{ind}}_{\mathrm{eq_{ah}}}(\mathbf{x}_{\mathrm{hp_{play}}}) = \frac{\mathbf{H}^{\mathrm{ind}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{e}}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}})}{\mathbf{H}^{\mathrm{ah}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}})} \cdot \frac{1}{\mathbf{H}_{\mathrm{ahm}} \cdot \mathbf{H}_{\mathrm{i_{ah}}} \cdot \mathbf{H}_{\mathrm{o_{hp}}} \cdot \mathbf{H}^{\mathrm{ind,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{e}}}(\mathbf{x}_{\mathrm{hp_{play}}})}. \tag{4.55}$$

Human Head Playback, Artificial Head Recording, and Probe Microphone Headphone Transfer Functions Relating equation 4.24 to equation 4.55 shows the human head equalization necessities for BS with AH recording dependent on probe microphone HPTFs. The results are formulated by

$$\begin{aligned}\mathbf{H}^{\mathrm{ind}}_{\mathrm{eq_{ah}}}(\mathbf{x}_{\mathrm{hp_{play}}}) = {} & \frac{1}{\mathbf{H}^{\mathrm{ind,h}}_{\mathrm{hptf_{pm}}}(\mathbf{x}_{\mathrm{hp_{hptf}}}, \mathbf{x}_{\mathrm{pm_{hptf}}})} \cdot \frac{\mathbf{H}^{\mathrm{ind}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{e}}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}})}{\mathbf{H}^{\mathrm{ah}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}})} \cdot \\ & \cdot \frac{\mathbf{H}^{\mathrm{ind,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{pm}}}(\mathbf{x}_{\mathrm{hp_{hptf}}}, \mathbf{x}_{\mathrm{pm_{hptf}}})}{\mathbf{H}^{\mathrm{ind,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{e}}}(\mathbf{x}_{\mathrm{hp_{play}}})} \cdot \frac{\mathbf{H}_{\mathrm{pm}} \cdot \mathbf{H}_{\mathrm{i_{pm}}}}{\mathbf{H}_{\mathrm{ahm}} \cdot \mathbf{H}_{\mathrm{i_{ah}}}}.\end{aligned} \tag{4.56}$$

According to equation 4.56, it is impossible to equalize BS with AH recording for human head playback using probe microphone HPTFs only. Errors may result from differences between the artificial and human head transfer characteristics and from different microphones and input equipment. Further, the probe microphone ear signal approximation and deviating HP positions during the HPTF measurement and playback situations can disrupt the equalization results.

Human Head Playback, Artificial Head Recording, and Blocked Auditory Canal Headphone Transfer Functions With equations 4.27 and 4.55, the equalization requirements for AH recording can be formulated in relation to blocked auditory canal HPTFs. For human head playback, this results in

$$\begin{aligned}\mathbf{H}^{\mathrm{ind}}_{\mathrm{eq_{ah}}}(\mathbf{x}_{\mathrm{hp_{play}}}) = {} & \frac{1}{\mathbf{H}^{\mathrm{ind,h,b}}_{\mathrm{hptf_{m}}}(\mathbf{x}_{\mathrm{hp_{hptf}}}, \mathbf{x}_{\mathrm{m_{hptf}}})} \cdot \frac{1}{\mathbf{H}^{\mathrm{ind,h}}_{\mathbf{p}_{\mathrm{m_b}},\mathbf{p}_{\mathrm{e}},\mathrm{hp}}(\mathbf{x}_{\mathrm{hp_{play}}}, \mathbf{x}_{\mathrm{hp_{hptf}}}, \mathbf{x}_{\mathrm{m_{hptf}}})} \cdot \\ & \cdot \frac{\mathbf{H}^{\mathrm{ind}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{e}}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}})}{\mathbf{H}^{\mathrm{ah}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}})} \cdot \frac{\mathbf{H}_{\mathrm{m}} \cdot \mathbf{H}_{\mathrm{i_{m}}}}{\mathbf{H}_{\mathrm{ahm}} \cdot \mathbf{H}_{\mathrm{i_{ah}}}}.\end{aligned} \tag{4.57}$$

Equation 4.57 shows that BS with AH recording cannot be equalized for human head playback using blocked auditory canal HPTFs. Errors may result from different sound pressures at the blocked auditory canal entrances and at the eardrums and from deviations between the human and artificial heads. Further deviations are possible due to varying HP positions during HPTF measurement and playback and due to different hardware.

Human Head Playback, Artificial Head Recording and Headphone Transfer Functions

To derive a relationship between the equalization requirements for AH recording and AH HPTFs for human head playback, equations 4.29 and 4.55 are combined. This results in

$$\mathbf{H}^{\mathrm{ind}}_{\mathrm{eq}_{\mathrm{ah}}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{play}}}) = \frac{1}{\mathbf{H}^{\mathrm{ah,h}}_{\mathrm{hptf}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{hptf}}})} \cdot \frac{\mathbf{H}^{\mathrm{ah,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{hptf}}})}{\mathbf{H}^{\mathrm{ind,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{e}}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{play}}})} \cdot \frac{\mathbf{H}^{\mathrm{ind}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{e}}}(\mathbf{x}_{\mathrm{h}_{\mathrm{ref}}}, \mathbf{x}_{\mathrm{ls}_{\mathrm{ref}}})}{\mathbf{H}^{\mathrm{ah}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{h}_{\mathrm{ref}}}, \mathbf{x}_{\mathrm{ls}_{\mathrm{ref}}})}. \qquad (4.58)$$

Equation 4.58 represents the typical situation when using AH recording in BS. If assumptions 6 and 8 are fulfilled, the synthesis equals the reference scene.

Assumption 8 (*Artificial Head Properties*)

> *The artificial head used for implementing a binaural synthesis system represents the individual listener's head regarding headphone and loudspeaker playback.*

Speaking descriptively, how well the binaurally synthesized ear signal match the reference scene ear signals depends on possibly occurring deviations of the HP positions during HPTF measurement and playback and on the similarity of the human and artificial heads. Only if the AH represents the human head regarding HP and LS playback, the synthesis results can be correct.

4.6.2 Equalization Requirements for Artificial Head Playback

AH playback represents an important tool in the design process of BS systems. For that reason, BS implemented for AH playback is included in addition to the human head situation. In this section, the equalization requirements are specified for each combination of the playback and recording situations. For all combinations, the equalization results achievable using the HPTFs introduced in section 4.4 are formulated.

Artificial Head Playback and Probe Microphone Recording

It is possible to implement BS based on probe microphone recording with the probe tube tips close to the eardrums. The equalization requirements for the AH playback of probe microphone recording based BS are given combining equations 4.37 and 4.43 by

$$\begin{aligned}\mathbf{H}^{\mathrm{ah}}_{\mathrm{eq}_{\mathrm{pm}}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{play}}}, \mathbf{x}_{\mathrm{pm}_{\mathrm{rec}}}) &= \\ &= \frac{1}{\mathbf{H}_{\mathrm{pm}} \cdot \mathbf{H}_{\mathrm{i}_{\mathrm{pm}}} \cdot \mathbf{H}_{\mathrm{o}_{\mathrm{hp}}} \cdot \mathbf{H}^{\mathrm{ah,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{play}}})} \cdot \frac{\mathbf{H}^{\mathrm{ah}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{h}_{\mathrm{ref}}}, \mathbf{x}_{\mathrm{ls}_{\mathrm{ref}}})}{\mathbf{H}^{\mathrm{ind}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{pm}}}(\mathbf{x}_{\mathrm{h}_{\mathrm{ref}}}, \mathbf{x}_{\mathrm{ls}_{\mathrm{ref}}}, \mathbf{x}_{\mathrm{pm}_{\mathrm{rec}}})}. \qquad (4.59)\end{aligned}$$

Artificial Head Playback, Probe Microphone Recording, and Headphone Transfer Functions With equations 4.24 and 4.59, the equalization requirements for the AH playback of BS with probe microphone recording can be given dependent on probe microphone HPTFs. The results are formulated by

$$\begin{aligned}\mathbf{H}^{\mathrm{ah}}_{\mathrm{eq}_{\mathrm{pm}}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{play}}},\mathbf{x}_{\mathrm{pm}_{\mathrm{rec}}}) = {} & \frac{1}{\mathbf{H}^{\mathrm{ind,h}}_{\mathrm{hptf}_{\mathrm{pm}}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{hptf}}},\mathbf{x}_{\mathrm{pm}_{\mathrm{hptf}}})}\cdot \\ & \cdot\frac{\mathbf{H}^{\mathrm{ind,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{pm}}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{hptf}}},\mathbf{x}_{\mathrm{pm}_{\mathrm{hptf}}})}{\mathbf{H}^{\mathrm{ah,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{play}}})}\cdot\frac{\mathbf{H}^{\mathrm{ah}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{h}_{\mathrm{ref}}},\mathbf{x}_{\mathrm{ls}_{\mathrm{ref}}})}{\mathbf{H}^{\mathrm{ind}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{pm}}}(\mathbf{x}_{\mathrm{h}_{\mathrm{ref}}},\mathbf{x}_{\mathrm{ls}_{\mathrm{ref}}},\mathbf{x}_{\mathrm{pm}_{\mathrm{rec}}})}.\end{aligned}\tag{4.60}$$

Equation 4.60 discloses that BS with probe microphone recording can be equalized for AH playback using probe microphone HPTFs if assumptions 4, 5, 6, and 8 hold true. In other words, the sound pressures at the probe microphone tube tips must represent the ear signals, the AH must equal the recording head and the HPTF measurement head for HP and LS reproduction, and the HP and probe microphone positions must be constant.

Artificial Head Playback, Probe Microphone Recording, and Blocked Auditory Canal Headphone Transfer Functions The equalization requirements for the AH playback of BS with probe microphone recording can be given dependent on blocked auditory canal HPTFs using equations 4.27 and 4.59. Hence, the requirements are formulated by

$$\begin{aligned}\mathbf{H}^{\mathrm{ah}}_{\mathrm{eq}_{\mathrm{pm}}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{play}}},\mathbf{x}_{\mathrm{pm}_{\mathrm{rec}}}) = {} & \frac{1}{\mathbf{H}^{\mathrm{ind,h,b}}_{\mathrm{hptf}_{\mathrm{m}}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{hptf}}},\mathbf{x}_{\mathrm{m}_{\mathrm{hptf}}})}\cdot\frac{\mathbf{H}^{\mathrm{ah}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{h}_{\mathrm{ref}}},\mathbf{x}_{\mathrm{ls}_{\mathrm{ref}}})}{\mathbf{H}^{\mathrm{ind}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{pm}}}(\mathbf{x}_{\mathrm{h}_{\mathrm{ref}}},\mathbf{x}_{\mathrm{ls}_{\mathrm{ref}}},\mathbf{x}_{\mathrm{pm}_{\mathrm{rec}}})}\cdot \\ & \cdot\frac{\mathbf{H}^{\mathrm{ind,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{e}}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{play}}})}{\mathbf{H}^{\mathrm{ah,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{play}}})}\cdot\frac{1}{\mathbf{H}^{\mathrm{ind,h}}_{\mathbf{p}_{\mathrm{m}_{\mathrm{b}}},\mathbf{p}_{\mathrm{e}},\mathrm{hp}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{play}}},\mathbf{x}_{\mathrm{hp}_{\mathrm{hptf}}},\mathbf{x}_{\mathrm{m}_{\mathrm{hptf}}})}\cdot\frac{\mathbf{H}_{\mathrm{m}}\cdot\mathbf{H}_{\mathrm{i}_{\mathrm{m}}}}{\mathbf{H}_{\mathrm{pm}}\cdot\mathbf{H}_{\mathrm{i}_{\mathrm{pm}}}}.\end{aligned}\tag{4.61}$$

Equation 4.61 reveals that blocked auditory canal HPTFs are not suited for the equalization of BS with probe microphone recording for AH playback. Errors result from different sound pressures in the blocked auditory canals and at the eardrums and differences between the artificial and human heads. Further errors may stem from different microphones and input systems, varying HP positions, and the probe microphone ear signal approximation.

Artificial Head Playback, Probe Microphone Recording, and Artificial Head Headphone Transfer Functions With equations 4.29 and 4.59, the equalization requirements for the AH playback of BS with probe microphone recording can be given in relation to AH HPTFs. The requirements are formulated by

$$\begin{aligned}\mathbf{H}^{\mathrm{ah}}_{\mathrm{eq}_{\mathrm{pm}}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{play}}},\mathbf{x}_{\mathrm{pm}_{\mathrm{rec}}}) = {} & \frac{1}{\mathbf{H}^{\mathrm{ah,h}}_{\mathrm{hptf}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{play}}})}\cdot \\ & \cdot\frac{\mathbf{H}^{\mathrm{ah}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{h}_{\mathrm{ref}}},\mathbf{x}_{\mathrm{ls}_{\mathrm{ref}}})}{\mathbf{H}^{\mathrm{ind}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{pm}}}(\mathbf{x}_{\mathrm{h}_{\mathrm{ref}}},\mathbf{x}_{\mathrm{ls}_{\mathrm{ref}}},\mathbf{x}_{\mathrm{pm}_{\mathrm{rec}}})}\cdot\frac{\mathbf{H}^{\mathrm{ah,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{hptf}}})}{\mathbf{H}^{\mathrm{ah,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{play}}})}\cdot\frac{\mathbf{H}_{\mathrm{ahm}}\cdot\mathbf{H}_{\mathrm{i}_{\mathrm{ah}}}}{\mathbf{H}_{\mathrm{pm}}\cdot\mathbf{H}_{\mathrm{i}_{\mathrm{pm}}}}.\end{aligned}\tag{4.62}$$

Equation 4.62 shows that BS with probe microphone recording can be equalized for AH playback using AH HPTFs if the same HP positions during the HPTF measurement and playback are ensured, if the probe microphone measurements represent the ear signals, and if no differences in the microphone and input system characteristics occur in the AH and probe microphone configurations. Summarizing, assumptions 4 and 6 must hold true, and the microphone and input system influences must be negligible.

This result is interesting insofar as the procedure can be used, combined with a setup allowing for HPTF measurement and playback with exact HP repositioning, to quantify the errors introduced by the probe microphone ear signal approximation. This procedure is exact for AH probe microphone recording and disturbed by differences between the human and artificial heads regarding LS reproduction when evaluating human head probe microphone recording.

Artificial Head Playback and Blocked Auditory Canal Recording

Using equations 4.39 and 4.43, the equalization requirements are specified to BS with blocked auditory canal recording and AH playback. This situation is formulated by

$$\begin{aligned} &\mathbf{H}^{\mathrm{ah}}_{\mathrm{eq}_{\mathrm{ah}}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{play}}}, \mathbf{x}_{\mathrm{m}_{\mathrm{rec}}}) = \\ &= \frac{\mathbf{H}^{\mathrm{ah}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{h}_{\mathrm{ref}}}, \mathbf{x}_{\mathrm{ls}_{\mathrm{ref}}})}{\mathbf{H}^{\mathrm{ind,b}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{m}}}(\mathbf{x}_{\mathrm{h}_{\mathrm{ref}}}, \mathbf{x}_{\mathrm{ls}_{\mathrm{ref}}}, \mathbf{x}_{\mathrm{m}_{\mathrm{rec}}})} \cdot \frac{1}{\mathbf{H}_{\mathrm{m}} \cdot \mathbf{H}_{\mathrm{i}_{\mathrm{m}}} \cdot \mathbf{H}_{\mathrm{o}_{\mathrm{hp}}} \cdot \mathbf{H}^{\mathrm{ah,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{play}}})}. \end{aligned} \tag{4.63}$$

Artificial Head Playback, Blocked Auditory Canal Recording, and Probe Microphone Headphone Transfer Functions With equations 4.24 and 4.63, the equalization required for BS with blocked auditory canal recording and AH playback is formulated dependent on probe microphone HPTFs. The result is given by

$$\begin{aligned} \mathbf{H}^{\mathrm{ah}}_{\mathrm{eq}_{\mathrm{ah}}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{play}}}, \mathbf{x}_{\mathrm{m}_{\mathrm{rec}}}) = &\frac{1}{\mathbf{H}^{\mathrm{ind,h}}_{\mathrm{hptf}_{\mathrm{pm}}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{hptf}}}, \mathbf{x}_{\mathrm{pm}_{\mathrm{hptf}}})} \cdot \frac{\mathbf{H}^{\mathrm{ind,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{pm}}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{hptf}}}, \mathbf{x}_{\mathrm{pm}_{\mathrm{hptf}}})}{\mathbf{H}^{\mathrm{ah,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{play}}})} \cdot \\ &\cdot \frac{\mathbf{H}^{\mathrm{ah}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{h}_{\mathrm{ref}}}, \mathbf{x}_{\mathrm{ls}_{\mathrm{ref}}})}{\mathbf{H}^{\mathrm{ind,b}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{m}}}(\mathbf{x}_{\mathrm{h}_{\mathrm{ref}}}, \mathbf{x}_{\mathrm{ls}_{\mathrm{ref}}}, \mathbf{x}_{\mathrm{m}_{\mathrm{rec}}})} \cdot \frac{\mathbf{H}_{\mathrm{pm}} \cdot \mathbf{H}_{\mathrm{i}_{\mathrm{pm}}}}{\mathbf{H}_{\mathrm{m}} \cdot \mathbf{H}_{\mathrm{i}_{\mathrm{m}}}}. \end{aligned} \tag{4.64}$$

Equation 4.64 reveals that the equalization of BS with blocked auditory canal entrance recording for AH playback is impossible with probe microphone HPTFs. The major error is introduced by the different sound pressures in the blocked auditory canals and at the eardrums. Further errors may result from deviations between the sound pressures at the probe microphone tube tips and the ear signals, deviations in the HP positions during HPTF measurement and playback, and differences between the AH and the recording head regarding HP reproduction as well as the AH and HPTF measurement head with regard to LS playback. In addition, deviations in the transfer characteristics of the probe and miniature microphones as well as of the input systems may disturb the results.

Artificial Head Playback, Blocked Auditory Canal Recording and Headphone Transfer Functions Comparing equations 4.27 and 4.63, it is possible to express the equalization requirements for BS with blocked auditory canal recording and AH playback in relation to blocked auditory canal HPTFs. The requirements are given by

$$\begin{aligned}\mathbf{H}^{\mathrm{ah}}_{\mathrm{eq_{ah}}}(\mathbf{x}_{\mathrm{hp_{play}}}, \mathbf{x}_{\mathrm{m_{rec}}}) = {} & \frac{1}{\mathbf{H}^{\mathrm{ind,h,b}}_{\mathrm{hptf_m}}(\mathbf{x}_{\mathrm{hp_{hptf}}}, \mathbf{x}_{\mathrm{m_{hptf}}})} \cdot \frac{\mathbf{H}^{\mathrm{ah}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}})}{\mathbf{H}^{\mathrm{ind,b}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{m}}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{m_{rec}}})} \cdot \\ & \cdot \frac{1}{\mathbf{H}^{\mathrm{ind,h}}_{\mathbf{p}_{\mathrm{m_b}},\mathbf{p}_{\mathrm{e}},\mathrm{hp}}(\mathbf{x}_{\mathrm{hp_{play}}}, \mathbf{x}_{\mathrm{hp_{hptf}}}, \mathbf{x}_{\mathrm{m_{hptf}}})} \cdot \frac{\mathbf{H}^{\mathrm{ind,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{e}}}(\mathbf{x}_{\mathrm{hp_{play}}})}{\mathbf{H}^{\mathrm{ah,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{hp_{play}}})} .\end{aligned} \tag{4.65}$$

Equation 4.65 shows that BS with blocked auditory canal entrance recording can be equalized for AH playback with blocked auditory canal HPTFs if assumptions 5, 6, 7 and 8 are fulfilled. These assumptions require identical miniature microphone positions in the recording situation and for the blocked auditory canal HPTF measurement. Further required are identical HP positions in the playback situation and during the HPTF measurement and identical blocking factors for HP presentation and LS presentation without HPs. Additionally, differences between the human and artificial head characteristics may cause erroneous results.

If the blocked auditory canal entrance recording and HPTF measurement are carried out using an AH, from a theoretical point of view representing a specific individual head, the following equalities hold, transferring the human to an artificial head situation:

$$\mathbf{H}^{\mathrm{ind,h,b}}_{\mathrm{hptf_m}}(\mathbf{x}_{\mathrm{hp_{hptf}}}, \mathbf{x}_{\mathrm{m_{hptf}}}) = \mathbf{H}^{\mathrm{ah,h,b}}_{\mathrm{hptf_{ahm}}}(\mathbf{x}_{\mathrm{hp_{hptf}}}) \tag{4.66}$$

$$\mathbf{H}^{\mathrm{ind,h}}_{\mathbf{p}_{\mathrm{m_b}},\mathbf{p}_{\mathrm{e}},\mathrm{hp}}(\mathbf{x}_{\mathrm{hp_{play}}}, \mathbf{x}_{\mathrm{hp_{hptf}}}, \mathbf{x}_{\mathrm{m_{hptf}}}) = \mathbf{H}^{\mathrm{ah,h}}_{\mathbf{p}_{\mathrm{m_b}},\mathbf{p}_{\mathrm{ahm_e}},\mathrm{hp}}(\mathbf{x}_{\mathrm{hp_{play}}}, \mathbf{x}_{\mathrm{hp_{hptf}}}, \mathbf{x}_{\mathrm{m_{hptf}}}) \tag{4.67}$$

$$\mathbf{H}^{\mathrm{ind}}_{\mathbf{p}_{\mathrm{m_b}},\mathbf{p}_{\mathrm{e}},\mathrm{ls}}(\mathbf{x}_{\mathrm{m_{rec}}}) = \mathbf{H}^{\mathrm{ah}}_{\mathbf{p}_{\mathrm{m_b}},\mathbf{p}_{\mathrm{ahm_e}},\mathrm{ls}}(\mathbf{x}_{\mathrm{m_{rec}}}) \tag{4.68}$$

$$\mathbf{H}^{\mathrm{ind,b}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{m}}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{m_{rec}}}) = \mathbf{H}^{\mathrm{ah,b}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{m}}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}) \tag{4.69}$$

$$\mathbf{H}^{\mathrm{ind,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{e}}}(\mathbf{x}_{\mathrm{hp_{play}}}) = \mathbf{H}^{\mathrm{ah,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{hp_{play}}}) \tag{4.70}$$

On this basis, the equalization requirements for BS with blocked auditory canal AH recording and HPTF measurement are given in addition to the same situation based on human head blocked auditory canal measurements. In the AH situation, equation 4.65 can be simplified using equation 4.52 to

$$\mathbf{H}^{\mathrm{ah}}_{\mathrm{eq_{ah}}}(\mathbf{x}_{\mathrm{hp_{play}}}) = \frac{1}{\mathbf{H}^{\mathrm{ah,h,b}}_{\mathrm{hptf_{ahm}}}(\mathbf{x}_{\mathrm{hp_{hptf}}})} \cdot \frac{\mathbf{H}^{\mathrm{ah}}_{\mathbf{p}_{\mathrm{m_b}},\mathbf{p}_{\mathrm{ahm_e}},\mathrm{ls}}(\mathbf{x}_{\mathrm{m_{rec}}})}{\mathbf{H}^{\mathrm{ah,h}}_{\mathbf{p}_{\mathrm{m_b}},\mathbf{p}_{\mathrm{ahm_e}},\mathrm{hp}}(\mathbf{x}_{\mathrm{hp_{play}}}, \mathbf{x}_{\mathrm{hp_{hptf}}}, \mathbf{x}_{\mathrm{m_{hptf}}})} . \tag{4.71}$$

The equalization requirements formulated by equation 4.71 show that BS with blocked auditory canal AH recording and AH playback can be equalized with blocked auditory canal AH HPTFs under two conditions. These conditions are equal HP positions during playback and HPTF measurement (assumption 6) and identical blocking factors for HP reproduction and LS playback without HPs (assumption 7).

Artificial Head Playback, Blocked Auditory Canal Recording, and Artificial Head Headphone Transfer Functions With equations 4.29 and 4.63, the equalization requirements for BS with blocked auditory canal entrance recording can be given for AH playback with respect to AH HPTFs. The resulting requirements are formulated by

$$\begin{aligned}\mathbf{H}^{\mathrm{ah}}_{\mathrm{eq}_{\mathrm{ah}}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{play}}},\mathbf{x}_{\mathrm{m}_{\mathrm{rec}}}) = {} & \frac{1}{\mathbf{H}^{\mathrm{ah,h}}_{\mathrm{hptf}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{play}}})} \cdot \frac{\mathbf{H}^{\mathrm{ah}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{h}_{\mathrm{ref}}},\mathbf{x}_{\mathrm{ls}_{\mathrm{ref}}})}{\mathbf{H}^{\mathrm{ind,b}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{m}}}(\mathbf{x}_{\mathrm{h}_{\mathrm{ref}}},\mathbf{x}_{\mathrm{ls}_{\mathrm{ref}}},\mathbf{x}_{\mathrm{m}_{\mathrm{rec}}})} \cdot \\ & \cdot \frac{\mathbf{H}^{\mathrm{ah,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{hptf}}})}{\mathbf{H}^{\mathrm{ah,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{play}}})} \cdot \frac{\mathbf{H}_{\mathrm{ahm}} \cdot \mathbf{H}_{\mathrm{i}_{\mathrm{ah}}}}{\mathbf{H}_{\mathrm{m}} \cdot \mathbf{H}_{\mathrm{i}_{\mathrm{m}}}}.\end{aligned} \tag{4.72}$$

Equation 4.72 shows that the equalization of a BS system implemented with blocked auditory canal entrance recording for AH playback is not possible using AH HPTFs only. The major error results from the different sound pressure signals at the blocked auditory canal entrances and at the eardrums. Further, differences in the HP positions during HPTF measurement and playback and different microphone and input system characteristics may introduce errors.

Artificial Head Playback and Recording

Using equations 4.40 and 4.43, it is possible to determine the equalization requirements for BS with AH recording for AH playback. The results are formulated by

$$\mathbf{H}^{\mathrm{ah}}_{\mathrm{eq}_{\mathrm{ah}}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{play}}}) = \frac{1}{\mathbf{H}_{\mathrm{ahm}} \cdot \mathbf{H}_{\mathrm{i}_{\mathrm{ah}}} \cdot \mathbf{H}_{\mathrm{o}_{\mathrm{hp}}} \cdot \mathbf{H}^{\mathrm{ah,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{play}}})}. \tag{4.73}$$

Artificial Head Playback and Recording, Probe Microphone Headphone Transfer Functions With equations 4.24 and 4.73, the equalization requirements for the AH playback of BS with AH recording can be given dependent on probe microphone HPTFs. These requirements are formulated in the frequency domain by

$$\begin{aligned}&\mathbf{H}^{\mathrm{ah}}_{\mathrm{eq}_{\mathrm{ah}}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{play}}}) = \\ &= \frac{1}{\mathbf{H}^{\mathrm{ind,h}}_{\mathrm{hptf}_{\mathrm{pm}}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{hptf}}},\mathbf{x}_{\mathrm{pm}_{\mathrm{hptf}}})} \cdot \frac{\mathbf{H}^{\mathrm{ind,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{pm}}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{hptf}}},\mathbf{x}_{\mathrm{pm}_{\mathrm{hptf}}})}{\mathbf{H}^{\mathrm{ah,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{play}}})} \cdot \frac{\mathbf{H}_{\mathrm{pm}} \cdot \mathbf{H}_{\mathrm{i}_{\mathrm{pm}}}}{\mathbf{H}_{\mathrm{ahm}} \cdot \mathbf{H}_{\mathrm{i}_{\mathrm{ah}}}}.\end{aligned} \tag{4.74}$$

Equation 4.74 reveals that BS implemented with AH recording and equalized for AH playback using probe microphone HPTFs shows no transmission errors if assumptions 4, 6, and 8 hold true. In other words, the human and artificial head characteristics must be identical with respect to LS reproduction, and the probe microphones and AH microphones including their input systems must show equal characteristics. It is further required that the sound pressure signals at the probe microphones represent the ear signals and that the HP positions during HPTF measurement and playback are identical.

Artificial Head Playback and Recording, Blocked Auditory Canal Headphone Transfer Functions For AH playback, the equalization requirements for BS based on AH recording are given in relation to blocked auditory canal HPTFs using equations 4.27 and 4.73. The requirements are in this case formulated by

$$\begin{aligned}\mathbf{H}_{\mathrm{eq}_{\mathrm{ah}}}^{\mathrm{ah}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{play}}}) = {} & \frac{1}{\mathbf{H}_{\mathrm{hptf}_{\mathrm{m}}}^{\mathrm{ind,h,b}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{hptf}}}, \mathbf{x}_{\mathrm{m}_{\mathrm{hptf}}})} \cdot \frac{1}{\mathbf{H}_{\mathbf{p}_{\mathrm{m}_{\mathrm{b}}},\mathbf{p}_{\mathrm{e}},\mathrm{hp}}^{\mathrm{ind,h}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{play}}}, \mathbf{x}_{\mathrm{hp}_{\mathrm{hptf}}}, \mathbf{x}_{\mathrm{m}_{\mathrm{hptf}}})} \cdot \\ & \cdot \frac{\mathbf{H}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{e}}}^{\mathrm{ind,h}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{play}}})}{\mathbf{H}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{ahm}}}^{\mathrm{ah,h}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{play}}})} \cdot \frac{\mathbf{H}_{\mathrm{m}} \cdot \mathbf{H}_{\mathrm{i}_{\mathrm{m}}}}{\mathbf{H}_{\mathrm{ahm}} \cdot \mathbf{H}_{\mathrm{i}_{\mathrm{ah}}}}.\end{aligned} \tag{4.75}$$

Equation 4.75 shows that the equalization of BS with AH recording and playback is not possible using blocked auditory canal HPTFs. Errors may result from different sound pressure signals at the blocked auditory canal entrances and at the eardrums in the playback situation, varying HP positions during HPTF measurement and playback, and from deviations of microphones and input systems.

Artificial Head Playback, Recording, and Headphone Transfer Functions Combining equations 4.29 and 4.73, it is possible to formulate the equalization requirements for BS with AH recording for AH playback dependent on AH HPTFs. The requirements for this situation, using an AH for every implementation step, are given by

$$\mathbf{H}_{\mathrm{eq}_{\mathrm{ah}}}^{\mathrm{ah}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{play}}}) = \frac{1}{\mathbf{H}_{\mathrm{hptf}_{\mathrm{ahm}}}^{\mathrm{ah,h}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{hptf}}})} \cdot \frac{\mathbf{H}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{ahm}}}^{\mathrm{ah,h}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{hptf}}})}{\mathbf{H}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{ahm}}}^{\mathrm{ah,h}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{play}}})}. \tag{4.76}$$

Equation 4.76 proves that BS with AH recording and playback can be equalized using AH HPTFs if the HP positions during HPTF measurement and playback are identical (assumption 6). This is the only situation that can be equalized using HPTFs solely.

4.7 Equalized Binaural Synthesis

In an actual BS playback situation, the audio signal to be presented is convolved with the BIRPs recorded according to section 4.3 and with equalization filters, as defined in section 4.6. The convolution products are in the situations discussed here presented to the subject by HPs, which is described mathematically in section 4.4. During the system development and evaluation processes, AH playback may be desirable since it is more reproducible and less time consuming than the human head evaluation (cf. section 5.2.1).

Using equations 4.12, 4.21, and 4.43, the ear signal spectra of the equalized BS situation can be given in their general form based on the introduced framework. Accordingly, the equalized BS produces the ear signal spectra

$$\begin{aligned}\mathbf{P}_{\mathrm{e}_{\mathrm{bs}}}^{\mathrm{ind,h}}(\mathbf{x}_{\mathrm{h}_{\mathrm{ref}}}, \mathbf{x}_{\mathrm{ls}_{\mathrm{ref}}}, \mathbf{x}_{\mathrm{mic}_{\mathrm{rec}}}, \mathbf{x}_{\mathrm{hp}_{\mathrm{play}}}, \mathbf{x}_{\mathrm{hp}_{\mathrm{hptf}}}, \mathbf{x}_{\mathrm{mic}_{\mathrm{hptf}}}) = {} & \mathbf{H}_{\mathrm{play}}^{\mathrm{ind,h}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{play}}}) \cdot \\ \cdot \mathbf{H}_{\mathrm{eq}}^{\mathrm{ind}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{play}}}, \mathbf{x}_{\mathrm{mic}_{\mathrm{rec}}}, \mathbf{x}_{\mathrm{hp}_{\mathrm{hptf}}}, \mathbf{x}_{\mathrm{mic}_{\mathrm{hptf}}}) \cdot \mathbf{H}_{\mathrm{rec}_{\mathrm{mic}}}^{\mathrm{ind}}(\mathbf{x}_{\mathrm{h}_{\mathrm{ref}}}, \mathbf{x}_{\mathrm{ls}_{\mathrm{ref}}}, \mathbf{x}_{\mathrm{mic}_{\mathrm{rec}}}) \cdot S_{\mathrm{ls}}.\end{aligned} \tag{4.77}$$

Based on equation 4.77, the equalized BS TFs are formulated independently of the recording procedure and HPTF measurement method in general by

$$\begin{aligned}
&\mathbf{H}_{\mathrm{bs}}^{\mathrm{ind,h}}(\mathbf{x}_{\mathrm{h}_{\mathrm{ref}}}, \mathbf{x}_{\mathrm{ls}_{\mathrm{ref}}}, \mathbf{x}_{\mathrm{mic}_{\mathrm{rec}}}, \mathbf{x}_{\mathrm{hp}_{\mathrm{play}}}, \mathbf{x}_{\mathrm{hp}_{\mathrm{hptf}}}, \mathbf{x}_{\mathrm{mic}_{\mathrm{hptf}}}) = \\
&= \frac{\mathbf{P}_{\mathrm{e_{bs}}}^{\mathrm{ind,h}}(\mathbf{x}_{\mathrm{h}_{\mathrm{ref}}}, \mathbf{x}_{\mathrm{ls}_{\mathrm{ref}}}, \mathbf{x}_{\mathrm{mic}_{\mathrm{rec}}}, \mathbf{x}_{\mathrm{hp}_{\mathrm{play}}}, \mathbf{x}_{\mathrm{hp}_{\mathrm{hptf}}}, \mathbf{x}_{\mathrm{mic}_{\mathrm{hptf}}})}{S_{\mathrm{ls}}} = \mathbf{H}_{\mathrm{play}}^{\mathrm{ind,h}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{play}}}) \cdot \\
&\cdot \mathbf{H}_{\mathrm{eq}}^{\mathrm{ind}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{play}}}, \mathbf{x}_{\mathrm{mic}_{\mathrm{rec}}}, \mathbf{x}_{\mathrm{hp}_{\mathrm{hptf}}}, \mathbf{x}_{\mathrm{mic}_{\mathrm{hptf}}}) \cdot \mathbf{H}_{\mathrm{rec}_{\mathrm{mic}}}^{\mathrm{ind}}(\mathbf{x}_{\mathrm{h}_{\mathrm{ref}}}, \mathbf{x}_{\mathrm{ls}_{\mathrm{ref}}}, \mathbf{x}_{\mathrm{mic}_{\mathrm{rec}}}).
\end{aligned} \tag{4.78}$$

Comparable to the reference scene evaluation in section 4.2, there are two ways of measuring the equalized BS ear signals: probe microphone measurement with the probe tube tips close to the eardrums, and AH measurement. Since only the AH measurement is possible without approximations imposed by the procedure (cf. section 5.1.4), it is selected for the BS evaluation discussed here. While not exactly representing human head playback, the AH evaluation allows assessing all system-theoretic aspects of BS without loss of generality (Völk 2010a). When transferring the results to human listeners, it is necessary to consider physical differences as for example the soft human tissue in contrast to the typically harder AH surface, varying eardrum impedances (Zwicker 1961a, Schmidt and Hudde 2009), or the frequently missing AH hair (Katz 2000, Treeby et al. 2007a,b).

Relating the spectra of the sequences representing the AH microphone outputs and the LS driving sequence in the evaluation situation results in the TFs

$$\begin{aligned}
&\mathbf{H}_{\mathrm{bs_{ver}}}^{\mathrm{ah,h}}(\mathbf{x}_{\mathrm{h}_{\mathrm{ref}}}, \mathbf{x}_{\mathrm{ls}_{\mathrm{ref}}}, \mathbf{x}_{\mathrm{mic}_{\mathrm{rec}}}, \mathbf{x}_{\mathrm{hp}_{\mathrm{play}}}, \mathbf{x}_{\mathrm{hp}_{\mathrm{hptf}}}, \mathbf{x}_{\mathrm{mic}_{\mathrm{hptf}}}) = \\
&= \frac{\mathbf{S}_{\mathrm{e_{bs}}}^{\mathrm{ah,h}}(\mathbf{x}_{\mathrm{h}_{\mathrm{ref}}}, \mathbf{x}_{\mathrm{ls}_{\mathrm{ref}}}, \mathbf{x}_{\mathrm{mic}_{\mathrm{rec}}}, \mathbf{x}_{\mathrm{hp}_{\mathrm{play}}}, \mathbf{x}_{\mathrm{hp}_{\mathrm{hptf}}}, \mathbf{x}_{\mathrm{mic}_{\mathrm{hptf}}})}{S_{\mathrm{ls}}},
\end{aligned} \tag{4.79}$$

describing the equalized AH BS situation. Figure 4.7 shows the AAA spectrogram of the AH approximation of a BS TF according to equation 4.79 for the synthesis[3] of the reverberant listening environment also represented by its reference scene in figure 4.2.

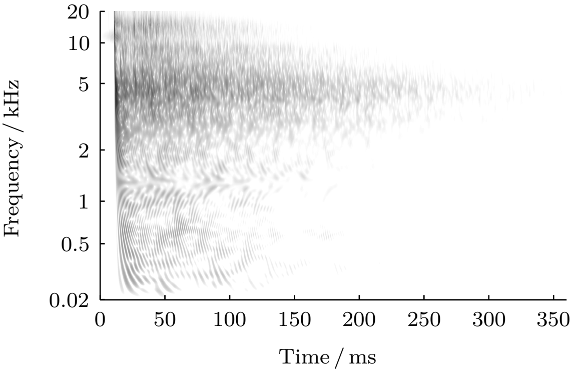

Figure 4.7: Auditory-adapted analysis spectrogram with 60 dB visible dynamic of the binaural synthesis of a loudspeaker in reverberant environment represented by means of its reference scene in figure 4.2. The synthesis was implemented employing artificial head recording and headphone transfer function measurement. The validation was carried out with the recording head.

[3] Sennheiser HD 800 headphones, Klein + Hummel Studio Monitor Loudspeaker O 98

The BS represented by figure 4.7 is implemented based on AH recording and HPTFs according to equation 4.76, with constant HP positions for HPTF measurement and validation. Figure 4.7 shows the BS of the reference scene depicted by figure 4.2, implemented and evaluated using AH$_\mathrm{c}$, with the microphone at the eardrum position. Therefore, a qualitative valuation of the BS result is possible by visually comparing figures 4.2 and 4.7, revealing good accordance of both ear signals. However, to address and compare the performance of BS systems, a quantitative quality criterion is necessary.

4.8 Binaural Synthesis Quality Criterion

The exemplary BS shown by figure 4.7 resembles the reference scene given by figure 4.2 qualitatively. In order to achieve a quantitative accuracy measure for binaurally synthesized ear signals, an analytic criterion referred to as binaural synthesis quality criterion (BSQC) is derived based on AH validation measurements in this section.

A BS system generates the reference scene ear signals if the BS TFs according to equation 4.78 equal the reference scene TFs defined by equation 4.3. For a BS system implemented using AH recording combined with AH HPTFs and validated by AH measurement, this is formulated mathematically. Combining the AH reference scene description formulated by equation 4.11 and the recording situation TFs given by equation 4.12 with the BS TFs represented by equation 4.79, the requirements for correct BS are given by

$$\begin{aligned}\mathbf{H}^{\mathrm{ah,h}}_{\mathrm{bs}_\mathrm{ver}}(\mathbf{x}_{\mathrm{h}_\mathrm{ref}},\mathbf{x}_{\mathrm{ls}_\mathrm{ref}},\mathbf{x}_{\mathrm{mic}_\mathrm{rec}},\mathbf{x}_{\mathrm{hp}_\mathrm{play}},\mathbf{x}_{\mathrm{hp}_\mathrm{hptf}},\mathbf{x}_{\mathrm{mic}_\mathrm{hptf}}) \overset{!}{=}& \\ \overset{!}{=} \mathbf{H}^{\mathrm{ah}}_{\mathrm{ref}_\mathrm{ahm}}(\mathbf{x}_{\mathrm{h}_\mathrm{ref}},\mathbf{x}_{\mathrm{ls}_\mathrm{ref}}) \cdot \mathbf{H}_{\mathrm{i}_\mathrm{ah}} \cdot \mathbf{H}_{\mathrm{ahm}} = \mathbf{H}^{\mathrm{ah}}_{\mathrm{rec}_\mathrm{ahm}}(\mathbf{x}_{\mathrm{h}_\mathrm{ref}},\mathbf{x}_{\mathrm{ls}_\mathrm{ref}}).&\end{aligned} \tag{4.80}$$

Equation 4.80 shows that the BS and the recording situation TFs acquired with the same AH must be frequency independently identical for the binaurally synthesized ear signals to equal the reference scene ear signals. This is defined as the BSQC (subscript *bsqc*)

$$\begin{aligned}&\mathbf{H}^{\mathrm{ah}}_{\mathrm{bsqc}}(\mathbf{x}_{\mathrm{mic}_\mathrm{rec}},\mathbf{x}_{\mathrm{hp}_\mathrm{play}},\mathbf{x}_{\mathrm{hp}_\mathrm{hptf}},\mathbf{x}_{\mathrm{mic}_\mathrm{hptf}}) = \\ &= \frac{\mathbf{H}^{\mathrm{ah}}_{\mathrm{rec}_\mathrm{ahm}}(\mathbf{x}_{\mathrm{h}_\mathrm{ref}},\mathbf{x}_{\mathrm{ls}_\mathrm{ref}})}{\mathbf{H}^{\mathrm{ah,h}}_{\mathrm{bs}_\mathrm{ver}}(\mathbf{x}_{\mathrm{h}_\mathrm{ref}},\mathbf{x}_{\mathrm{ls}_\mathrm{ref}},\mathbf{x}_{\mathrm{mic}_\mathrm{rec}},\mathbf{x}_{\mathrm{hp}_\mathrm{play}},\mathbf{x}_{\mathrm{hp}_\mathrm{hptf}},\mathbf{x}_{\mathrm{mic}_\mathrm{hptf}})} \overset{!}{=} \mathbf{1}.\end{aligned} \tag{4.81}$$

The BSQC is evaluated in the following for two exemplary BS systems aiming at simulating the situation depicted by figures 4.2 and 4.7. The first system is based on AH playback, recording, and HPTFs according to equation 4.76. The second system is based on AH playback, blocked auditory canal AH recording, and blocked auditory canal AH HPTFs according to equation 4.71. Both systems are implemented using AH$_\mathrm{c}$, supporting the positioning of the microphone at the blocked auditory canal entrance and at the eardrum location without removing the HPs. Further, a measurement sequence is selected that allows for approximately constant HP positions during HPTF measurement and validation and consequently results in approximately constant HP to AH microphone TFs

$$\mathbf{H}^{\mathrm{ah,h}}_{\mathbf{u}_\mathrm{hp},\mathbf{p}_\mathrm{ahm}}(\mathbf{x}_{\mathrm{hp}_\mathrm{hptf}}) \approx \mathbf{H}^{\mathrm{ah,h}}_{\mathbf{u}_\mathrm{hp},\mathbf{p}_\mathrm{ahm}}(\mathbf{x}_{\mathrm{hp}_\mathrm{play}}). \tag{4.82}$$

Additionally, the measurement sequence ensures identical microphone positions for BTFP recording, HPTF measurement, and playback with open and blocked auditory canals.

4.8.1 Open Auditory Canal Artificial Head Binaural Synthesis Quality Criterion

According to equation 4.76, the equalization of BS with AH recording and playback is possible using AH HPTFs if equation 4.82 holds true. Since $\mathrm{AH_c}$ and the selected measurement procedure are developed to fulfill equation 4.82, the BS with $\mathrm{AH_c}$ recording, HPTFs, and playback is expected to provide frequency independent transfer characteristics with regard to the reference scene. Figure 4.8 shows the corresponding BSQC according to equation 4.81, representing the resulting BS error. Thereby, HP specimen a) is indicated by the black contours and HP specimen b) by the gray contours[4].

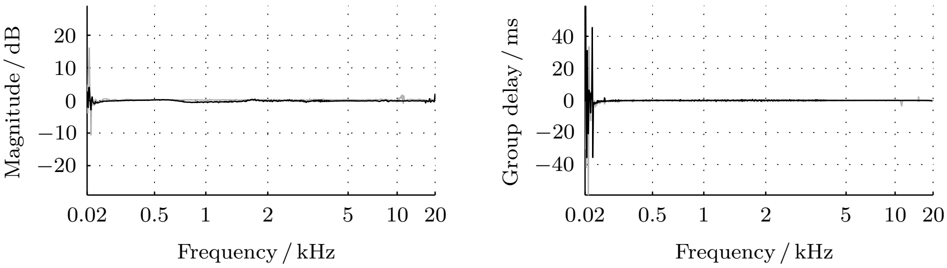

Figure 4.8: Error of the binaural synthesis of a loudspeaker in a reverberant environment as predicted by the binaural synthesis quality criterion. Reference scene identical to figure 4.2. Binaural synthesis implemented with artificial head recording and artificial head headphone transfer functions, validation at the eardrum of the same head. The black contours indicate headphone specimen a), the gray contours specimen b).

The BSQC reveals almost perfect, nearly identical synthesis for both HP specimens. Apart from artifacts in the frequency range below the transmission bandwidth of the reference scene LS (cf. figure 4.1), which cannot be attributed to the BS system, the magnitude spectra in the left panel and the group delays depicted on the right show spurious artifacts below the measurement accuracy. The global negative group delay shift of about 3 ms is introduced by the causal finite impulse response equalization filters (cf. section 5.4). Neglecting these methodical issues, the validity of the ear signals created by the BS systems implemented for AH playback with AH measurements is confirmed by the BSQC, as predicted by equations 4.76 and 4.82.

4.8.2 Blocked Auditory Canal Artificial Head Binaural Synthesis Quality Criterion

Based on equation 4.71, it is not necessarily clear whether BS with blocked auditory canal AH recording equalized by blocked auditory canal AH HPTFs provides frequency inde-

[4] a) Sennheiser HD 800, b) Stax λ pro NEW headphones

pendent transfer characteristics for standard AH playback. To evaluate this configuration, figure 4.9 shows the BSQC according to equation 4.81, implemented with AH_c. Again, HP specimen a) is indicated by the black contours and HP specimen b) by the gray contours.

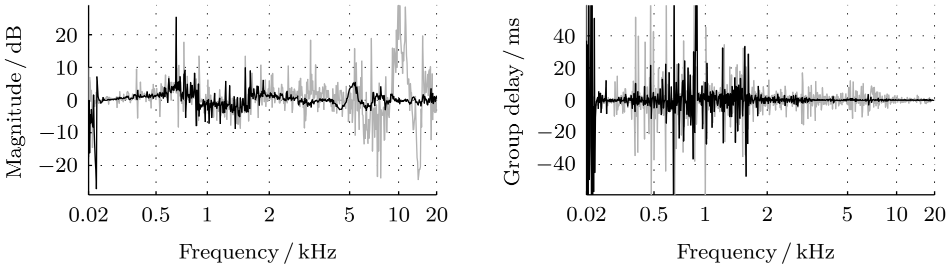

Figure 4.9: Error of the binaural synthesis of a loudspeaker in a reverberant environment as predicted by the binaural synthesis quality criterion. Reference scene identical to figure 4.2. Binaural synthesis implemented with blocked auditory canal artificial head recording and headphone transfer functions, validation at the eardrum of the same head. The black contours indicate headphone specimen a), the gray contours specimen b).

The BSQC reveals frequency dependence of magnitude spectrum and group delay for both HPs, in addition to the methodical artifacts also visible in figure 4.8. It is further noticeable that HP specimen can influence the BS transfer characteristics. The partial error of the BS procedure according to equation 4.71 not accounted for by the equalization with HPTFs only depends on the HP position in the recording situation. It is not possible to demonstrate based on a single measurement whether the setup as such or in combination with the specific recording situation HP position causes the observed frequency dependence. In order to address this question, figure 4.10 shows the results of seven measurements with HP specimen b).

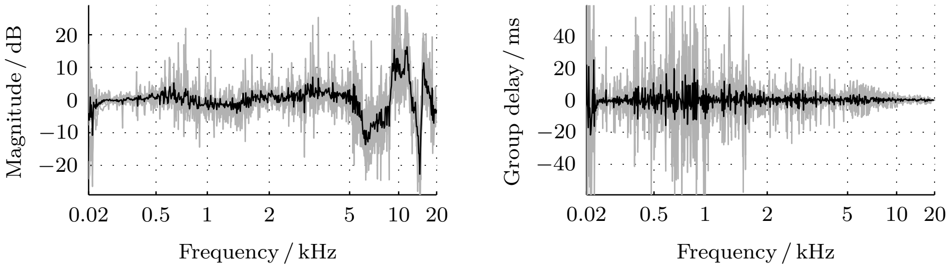

Figure 4.10: Error of the binaural synthesis of a loudspeaker in a reverberant environment as predicted by the binaural synthesis quality criterion. Binaural synthesis implemented with blocked auditory canal artificial head recording and headphone transfer functions, validation with the same head. Seven measurements for different sound incidence directions (gray contours) and averages (black contours), headphone specimen b).

Depicted in figure 4.10 are results for different sound incidence directions with intermediate HP repositioning (gray) and the corresponding averages of magnitude spectra and group delays (black). The global magnitude spectrum characteristics are visible in all measurements and may therefore be attributed to the prototypical measurement setup used. While the spurious group delay artifacts appear less pronounced on average, no obvious global group delay structure arises. It can be concluded that the global structure of the magnitude spectrum in figure 4.9 is not influenced by a specific recording situation HP position. This conclusion confirms the influence of the HP specimen on the transfer characteristics of BS with blocked auditory canal recording since the single difference between the BS systems represented by figure 4.9 is the HP employed.

Consequently, the selection of appropriate HPs respectively the equalization of the remaining error of equation 4.71 represent a crucial implementation factor for BS with miniature microphone recording at the blocked auditory canal entrance, especially regarding the frequency independence of the resulting magnitude spectrum. It is not necessarily clear that the equalization of the error remaining after the HPTF based equalization is possible because the symbolically defined blocking factors, connecting the open and blocked auditory canal situations and defining the error remaining in equation 4.71, relate two different systems. If the frequency dependencies visible in figures 4.9 and 4.10 are caused by resonances in the HP-head system, equalization may not be possible. However, if HPs can be found that allow for a frequency independent BSQC, which appears possible based on the data indicated by the black contours in figure 4.9, no further equalization would be necessary. The definition of appropriate HPs and an HP selection criterion are introduced in section 5.3 (cf. Völk 2010a, 2012a).

4.9 Summary

Based on the system-theoretic framework introduced in this chapter, recommendations for the design of binaural synthesis systems can be given. Since this chapter deals with the theoretic background, not with implementation details, it is based on the best case situation insofar as all necessary prerequisites and assumptions are identified and marked, but neither is their feasibility verified, nor are the consequences addressed that arise if an assumption is not valid. It is likely for certain implementation factors, as for example the reproducibility of the headphone position for headphone transfer function measurement and playback or constraints regarding the practical implementation of individual equalization filters, to influence the analysis given above. These factors have to be accounted for in addition to the best case situation discussed in the present chapter and are therefore addressed, based on the assumptions identified in this chapter, in chapter 5. However, since it provides the theoretical and methodical basis of binaural synthesis, the best case scenario introduced and discussed in the present chapter represents the baseline situation for further considerations.

Three recording methods for the binaural transfer function pairs required for binaural synthesis are addressed in this chapter based on a system-theoretic discussion of the partial systems involved in the binaural synthesis process: probe microphone recording with the probe tube tips close to the eardrums, miniature microphone recording at the entrances to the blocked auditory canals, and artificial head recording. Regarding the binaural synthesis playback situation, the playback to artificial heads is considered in addition to the usual human head playback. The artificial head reproduction is included since it allows for simple and reproducible measurements, especially in the design and implementation processes of binaural synthesis. All possible combinations of the partial systems are discussed with the aim of formulating the respective error system theoretically for all possibly occurring combinations, which is considered relevant since incorrect procedures are employed in practical applications where aspects as for example cost, effort, availability of binaural transfer function pairs, and implementation time may overrule the necessity of the theoretically optimal solution. Facing the effort of recording binaural impulse response pairs, application designers and researchers may decide to work with theoretically suboptimal systems (e. g. Keyrouz and Diepold 2008, Habigt et al. 2010, Jürgens and Werner 2012). It is important to take the incorrect stimulus into account when interpreting the results of listening experiments conducted using suboptimal binaural synthesis systems. Therefore, the system-theoretical framework derived in this chapter covers, for the discussed recording methods, all scenarios that may arise if binaural synthesis system designers combine previously acquired or publicly available system components, regardless of the correctness of the achievable result.

For providing a tool to indicate the authenticity of binaurally synthesized ear signals, a binaural synthesis quality criterion is introduced based on the artificial head validation of the binaural synthesis transfer functions. In combination with the system-theoretic framework derived, the binaural synthesis quality criterion can be used to identify errors in the synthesis procedure and to track them down to the causative partial system.

In general, correct binaural synthesis is possible only using recordings and headphone transfer functions of the subject the synthesis is intended for. This situation is referred to as individual binaural synthesis. It is also possible to employ a different subject or an artificial head for recording and headphone transfer function measurement, which is denoted nonindividual synthesis. The perceptual consequences of nonindividual measurements must be assessed by listening experiments as described in chapter 5 since the physical measurement of human head ear signals is not possible in a strict sense. In the following, summaries of the major findings of the present chapter regarding individual binaural synthesis are given and discussed in view of implementations separately for each recording method covered by the system-theoretic framework.

Binaural Synthesis with Probe Microphone Recording Developing binaural synthesis is possible based on probe microphone recording and headphone transfer function measurement with the probe tubes in the auditory canals, close to the eardrums. The resulting ear signals equal the reference scene ear signals to the degree at which the sound pressures at the probe microphones represent the ear signals, assuming identical microphone positions during recording and headphone transfer function measurement and equal headphone positions for playback and headphone transfer function measurement (cf. equation 4.45). However, to protect the listeners from injuries, probe microphone recording with the tube tips in the auditory canals, close to the eardrums, requires a careful and time-consuming procedure, ideally carried out by trained and accredited audiologists. Further, dependent on the headphones employed, the probe tubes may cause headphone leakage effects and therefore erroneous results of the headphone transfer function measurements.

Binaural Synthesis with Blocked Auditory Canal Recording With binaural synthesis based on miniature microphone recording and headphone transfer function measurement at the entrances to the blocked auditory canals, the reference scene ear signals can be recreated completely under three conditions (equation 4.53): the headphone transfer functions must be measured at the recording microphone positions, the headphone positions must be identical for headphone transfer function measurement and playback, and remaining errors of the binaural synthesis transfer functions must be accounted for, preferably by selecting appropriate headphones. Procedures to account for possibly remaining errors, including the definition and selection of appropriate headphones, are discussed in section 5.3. Carrying out miniature microphone measurements at the blocked auditory canal entrances is possible for non-trained personnel and typically requires less time than probe microphone measurements at the eardrums. Furthermore, blocked auditory canal entrance measurements contain less individual information than probe microphone eardrum measurements, which is advantageous for nonindividual binaural synthesis. Usually, it is possible to position the miniature microphone cables without causing leakage effects when measuring headphone transfer functions. For these reasons, blocked auditory canal entrance recording is regarded as the approach providing the highest potential of creating correct individual ear signals by binaural synthesis while introducing the least errors with nonindividual procedures and is therefore considered the preferable approach to binaural synthesis.

Binaural Synthesis with Artificial Head Recording If the probe microphone headphone transfer functions of the actual listener are used to equalize binaural synthesis implemented with artificial head recording, the synthesized ear signals may deviate from the intended ear signals due to physical differences between the artificial and human heads and due to artifacts of the probe microphone ear signal approximation (equation 4.56). Other possible sources of errors can include differences between the microphones and the input equipment of the recording and headphone transfer function measurement situations, and deviating headphone positions during playback and headphone transfer function measurement. In case using probe microphone headphone transfer functions is not justified or possible, artificial head headphone transfer functions may be employed (cf. equation 4.58). In this situation, deviations between the individual and artificial head headphone transfer functions may introduce additional errors of the synthesized ear signals. Blocked auditory canal artificial head recording equals, from a system-theoretic point of view, nonindividual blocked auditory canal recording with a specific, in this case artificial, subject.

5 Practical Aspects of Applied Binaural Synthesis

Binaural synthesis (BS) as a physically motivated virtual acoustics method using headphone (HP) reproduction attempts to synthesize the sound pressure signals at the eardrums, the so-called ear signals. The BS procedure is revised system theoretically in chapter 4, where assumptions allowing for the theoretical derivation are identified. According to section 4.9, the preferable approach to BS is based on the actual listener's binaural impulse response pairs (BIRPs) and headphone transfer functions (HPTFs). This approach is referred to as individual BS, a combination of individual recording and individual HPTF measurement, in contrast to the respective nonindividual procedures (definition 27).

Definition 27 (*Individual and Nonindividual Binaural Synthesis*)

> *Individual recording and individual headphone transfer function measurement for binaural synthesis is carried out with the actual listener, in contrast to the respective nonindividual procedures. Individual or nonindividual binaural synthesis implements both processes accordingly, while the two remaining combinations are referred to as mixed binaural synthesis.*

In this chapter, the *applicability and validity of the assumptions* made in chapter 4 are addressed with respect to individual, nonindividual, and mixed BS. As the reference scene is assumed static in chapter 4, the first section of the present chapter includes an overview of the theory and digital implementation of dynamic BS. Further, this section covers deviations between reference scene and recording situation, which are deviations from assumptions 1, 4, and 8. Statistics of the amount of typical deviations are given and methods to deal with resulting shortcomings are introduced and discussed with regard to physical and perceptual consequences.

The second section addresses discrepancies between either the recording and HPTF measurement situations or the playback and HPTF measurement situations, which are violations of assumptions 3, 5, and 6. These discrepancies may be summarized as *aspects of the BS equalization procedure*, discussed here regarding physical consequences, while perceptual aspects are addressed in the fourth section of this chapter.

In the third section, the necessity of selecting the HPs to be used in BS with blocked auditory canal entrance recording is shown and a *headphone selection criterion* based on artificial head (AH) measurements is introduced. This criterion in principle checks the validity of assumption 7, that is the equality of the so-called blocking factors, the sound pressure transformations between the eardrum and the entrance to the blocked auditory canal for HP reproduction and loudspeaker (LS) playback without HPs.

The fourth section reconsiders and clarifies the role of the ear signals in HP reproduction, which has been controversially discussed in the literature (Rudmose 1982, Fastl et al. 1985). Thereby, the *validity of the BS procedure* is proven according to assumption 2, the hypothesis of the ear signal recreation representing a sufficient BS design goal. This section may be regarded as a summarizing perceptual verification of the physical BS aspects discussed in the preceding, finalized by a reflection on implications of the results on the modeling of loudness and localization. A summary concludes the chapter.

5.1 Deviations between Reference Scene and Recording Situation

The BS reference scene definition selected for this thesis (definition 25) assumes an idealized static situation as the synthesis target (assumption 1). If a real situation has to be synthesized, this assumption is usually not fulfilled since listeners, sound sources, and the environment are subject to time variance. Thus, aspects of dynamic BS are discussed in the following. Typical approaches of modeling time-variant situations require spatial discretization. Therefore, issues regarding this discretization are pointed out with a special focus on its perceptual consequences. Implications of different spatially discrete recording procedures, especially with reduced complexity compared to individual recording, as for example AH or nonindividual recording, conclude this section.

5.1.1 Time Variance

If temporal scene variations are to be considered in BS, the question arises as to how far the situation must change before requiring a reaction of the BS system. Physically strict, it is necessary to account for every change in the situation. In implementations, the time it takes a BS system to adapt to a change in the situation is limited by several factors identified in the following. The discussion is given specifically for digital BS implementations, representing the almost exclusive situation nowadays, while the main concept also applies to analog systems.

Audio Output Latency For a digital system working at the sample rate $f_\mathrm{s} = 1/T_\mathrm{s}$, the earliest reaction is possible at the sampling instant following the respective notification. Consequently, the minimum latency a digital signal processing system can guarantee equals T_s. Typical audio interfaces and driver architectures split the audio signal into blocks of K samples and carry out the processing block-wise (Steinberg Media Technologies GmbH 2006, Apple Inc. 2007, Microsoft Corp. 2009, Iwai 2009). This procedure increases the average system reaction time since the presently processed block is typically locked and no longer accessible for modifications. High quality audio interfaces, as regarded here, implement double-buffer systems, with one buffer accessible to the application while the content of the other buffer is played back (cf. Steinberg Media Technologies GmbH 2006). To ensure playback without dropouts, the currently processed block is handed over to the interface driver with a device specific overhead of K' samples before the playback is finished. If the notification requesting a system state change occurs right after a block is transferred to the interface and after the subsequent block is locked, the system reaction is delayed further by $K' + 2K$ samples. Consequently,

$$T'_\mathrm{d}\left(f_\mathrm{s}, K, K'\right) = T_\mathrm{s} \cdot \left(1 + 2K + K'\right) \tag{5.1}$$

represents the worst case reaction time of a block based digital signal processing system after the triggering notification. In this case, processing times $T_{\mathrm{p}} < T_{\mathrm{s}} \cdot K = T'_{\mathrm{p}}$ per block introduce no further delay. Typical block sizes of audio signal processing systems correspond to latencies in the range from 1 ms to 50 ms (RME – Intelligent Audio Solutions 2011). Therefore, the delay a high quality digital audio signal processing system can guarantee lies, depending on the specific hardware, at some 10 ms. A frequently used setting with a block size $K = 256$, an overhead $K' = 50$, and the sample rate $f_{\mathrm{s}} = 44.1\,\mathrm{kHz}$ provides, according to equation 5.1, the worst case latency $T'_{\mathrm{d}}\,(44.1\,\mathrm{kHz}, 256, 50) \approx 13\,\mathrm{ms}$.

Overall Binaural Synthesis Latency Additional implementation specific delays have to be considered in BS, for example the delivery time T_{n} of the notification requesting a system reaction and the processing time T''_{p} required in addition to $T'_{\mathrm{p}} = T_{\mathrm{s}} \cdot K$ for providing the updated signal. Assuming a typical system with no further delays, the worst case BS latency is given by

$$T_{\mathrm{d}}\,(f_{\mathrm{s}}, K, K') = T_{\mathrm{n}} + T''_{\mathrm{p}} + T'_{\mathrm{d}}\,(f_{\mathrm{s}}, K, K')\,. \tag{5.2}$$

The just noticeable BS latency, that is the just noticeable delay of the ear signal adaptation with respect to the corresponding situation change, has been addressed by different studies. Sandvad (1996) found a significant increase of the intra-individual variation of horizontal and vertical free-field localization results when increasing the BS latency above 96 ms. However, the intra-individual variation of the results in Sandvad's optimal configuration exceeds the reference scene variation, indicating a non-ideal BS system. Brungart et al. (2005) report different BS latency detection thresholds measured with and without a direct comparison to the reference scene. Without the direct comparison, lowest detectable latencies of some 60 ms were found (supported also by the data of Felderhoff et al. 1998), while the latency thresholds with direct comparison are estimated to about 35 ms. Assuming the lower value reported by Brungart et al. as the absolute threshold,

$$T_{\mathrm{d}}\,(f_{\mathrm{s}}, K, K') = T_{\mathrm{n}} + T''_{\mathrm{p}} + T'_{\mathrm{d}}\,(f_{\mathrm{s}}, K, K') \overset{!}{\leq} 35\,\mathrm{ms} \tag{5.3}$$

represents the worst case latency requirement for transparent BS. Based on equation 5.3, the maximum signal processing time for transparent BS is given by

$$T_{\mathrm{p}} = T'_{\mathrm{p}} + T''_{\mathrm{p}} \overset{!}{\leq} T'_{\mathrm{p}} + 35\,\mathrm{ms} - T_{\mathrm{n}} - T'_{\mathrm{d}}\,(f_{\mathrm{s}}, K, K') = T_{\mathrm{p}_{\mathrm{max}}}\,. \tag{5.4}$$

Tracking systems frequently used for the notification in BS provide latencies in the range of $T_{\mathrm{n}} \approx 13\,\mathrm{ms}$ (Polhemus 2005), typical audio output latencies lie at $T'_{\mathrm{d}}\,(44.1\,\mathrm{kHz}, 256, 50) \approx 13\,\mathrm{ms}$. Consequently, the signal processing latency for a transparent dynamic BS system must not exceed $T_{\mathrm{p}_{\mathrm{max}}} \approx 15\,\mathrm{ms}$, which requires elaborate signal processing routines, especially for the BS of multi-source or reverberant environments.

Piecewise Time-Invariant System Modeling For incorporating time variance in BS, piecewise time-invariant system modeling can be employed (Wenzel 1995, Inanaga et al.

1995, Horbach 1997, Gardner 1997). Thereby, the linear dynamic system to be synthesized is idealized and assumed time-invariant within specified temporal intervals and described by an array of impulse responses (IRs) representing the system states of interest. Ideally, IRs are measured or modeled for every possible system state. If this effort is not justified or feasible, specific states are selected, sampling the continuum of system states. Whether the discretization causes information reduction depends on the situation to be simulated and the implementation (Strauss 1998, cf. section 5.1.2).

The discretization due to piecewise time-invariant system modeling in the context of BS must not be confused with a spatial sound field sampling, where the spatial sampling theorem applies (Spors 2005). In contrast to spatial sampling as for example in wave field synthesis implementations (cf. section 3.1.2), the IRs are applied in a BS playback situation sequentially, not simultaneously, employing adaptive signal processing for selecting the BIRP representing the current situation best. Consequently, no spectral alias artifacts occur in each static situation, but the degree of authenticity of the dynamic synthesis depends on the grid resolution and therefore on frequency.

Ideally, the IR update process is carried out transparently, that is inaudibly (cf. definition 19). As discussed above, transparent discrete IR adaptation is possible for overall BS delays below about 35 ms. Different algorithms and parameter sets have been developed to implement the IR update in BS (Gardner 1995, Torger and Farina 2001, Völk et al. 2007). Typical approaches are adapted to the block-wise audio interface signal processing by updating one block of a temporally partitioned IR every time the interface driver calls a new signal block. This temporal partitioning of signal and IRs requires an overlap-add or overlap-save procedure to ensure the correct convolution (Zölzer 2005, pp. 171–177). Depending on the time variance modeling and the situation to be simulated, IR updates are possible either after all partitions of the current IR have been processed or every time a signal block is handed to the interface driver. The actual update may be carried out exchanging or interpolating the current and the target IRs.

Perceptive Aspects of Dynamic Binaural Synthesis Wallach (1939, 1940) showed that source movements can decrease the number of front-back confusions in auditory localization and expected a similar effect also due to head movements. This expectation appears plausible assuming the hearing system exploits ear signal variations induced by changes of the geometric relations of head and sound source positions. Technically, such an exploitation is possible only if the geometric relations, especially the directions of the position changes, are known to the listener. Wallach's hypothesis is validated by Thurlow and Runge (1967), Wightman and Kistler (1999), and Iwaya et al. (2003), who report a reduction of front-back confusions due to head movements and listener induced source movements, but no reduction due to experimenter induced source movements. Thurlow and Runge further found increasing localization accuracy due to head movements.

For BS, Begault et al. (2001) report significantly less front-back confusions with a dynamic versus the corresponding static system[1]. Regarding localization accuracy or externalization, Begault et al. found no effect of dynamic versus static BS, supported

[1] Sennheiser HD 430 headphones, Polhemus 3 Space FasTrack, speech excerpts

with dynamic BS[2] by Iwaya et al. (2004). However, lacking correct BS equalization according to chapter 4, the results of Begault et al. (2001) and Iwaya et al. (2004), which contradict the data reported for real sources by Thurlow and Runge (1967), may contain systematic errors. Regarding loudness transfer, experiments conducted in the course of this thesis indicate negligible differences of the results acquired with static versus dynamic BS (section 5.4, Völk et al. 2011d, Völk and Fastl 2011a).

5.1.2 Discrete Measurement Grid

The BIRP selection in dynamic BS is typically controlled based on a geometric scene description, material and source properties of the scene, and the position and orientation of the listener, usually determined by a head-tracking system (Horbach et al. 1999). Regardless whether the BIRPs are acquired by measurement or simulation (data based or model based, cf. definition 24), spatial continuous scene adaptation is not possible in dynamic BS as described, and a discrete spatial grid must be introduced.

Within a three-dimensional scenario, all objects are located with six degrees of freedom, three translational and three rotational. Considering a static listening environment with the listener representing the only moving element, all listening situations are at a specific instant of time, that is for a section of the block-wise processed audio signal, unambiguously identified by the listener position and orientation. For dynamically varying scenarios, the positions and orientations of all objects within the scene must be taken into account. In general, every static situation is described by as many BIRPs as there are sound sources.

In the following, BS is discussed separately for anechoic and reverberant listening environments. While anechoic conditions can be regarded as a degenerated reverberant situation without reverberation, the synthesis of reverberant environments is not covered by the theory for anechoic conditions. However, the anechoic situation is discussed separately since it allows for reducing the required effort and has been used in earlier studies on the directional resolution of BS (e. g. Hoffmann and Møller 2005b).

Anechoic Conditions In an anechoic listening environment, the relative geometric arrangement of sources and listener describes a situation completely. For that reason, the virtual sources are positioned for the BS of anechoic listening situations with the desired orientation at the correct distance and angle relative to the center of the listener's head. Consequently, if for M_s different sources M_d distances, M_so source orientations, and M_ho listener orientations are to be included, $M_\mathrm{ir,ae} = M_\mathrm{s} \cdot M_\mathrm{d} \cdot M_\mathrm{so} \cdot M_\mathrm{ho}$ sets of BIRPs are necessary. In case the application scenario allows for a reduced number of distances or orientations per source, the number of BIRPs can be reduced accordingly. The maximum amount is given assuming the same numbers of distances and orientations per source.

The BIRP update for source positioning relative to the listener's head is implemented by transforming the source location to the momentary head-related coordinates and selecting the BIRP matching the angle relative to the head center, the listener and source orientations, and the source distance best. This procedure is according to definition 28

[2] STAX SR-202, Polhemus 3 Space FasTrack, individual and nonindividual synthesis

referred to as head-related BS. The spatial resolution of head-related BS and possible source positions are determined by the available BIRPs.

Reverberant Environments For the BS of reverberant listening environments, the absolute positions of listener and sources within the scenario must be taken into account. The maximum number of BIRPs is computed by $M_{\mathrm{ir}} = M_{\mathrm{s}} \cdot M_{\mathrm{spo}} \cdot M_{\mathrm{hpo}} \cdot M_{\mathrm{so}} \cdot M_{\mathrm{ho}}$, assuming the same number of listener and source orientations for each of M_{s} sources, with M_{spo} source and M_{hpo} head positions. The BIRP update for the BS of reverberant environments is implemented transforming the momentary head orientation and position to world coordinates and picking the BIRP matching the resulting situation best. According to definition 28, this procedure is referred to as room-related BS.

Definition 28 (*Head- and Room-Related Binaural Synthesis*)

> *Dynamic binaural synthesis positioning the virtual sources relative to the listener's head is referred to as head-related binaural synthesis, whereas the sources are positioned in a fixed world coordinate system in room-related binaural synthesis.*

Consequently, the statement *"In binaural synthesis virtual sound sources are implemented with locations relative to a listener"* (Minnaar et al. 2005) holds true for head-related, but not for room-related BS. The resolution of room-related BS is given by the available BIRPs, while possible source positions are, in contrast to head-related BS, determined by the source positions where BIRPs are available.

Simplified Configurations In order to reduce the effort of implementing a fully three-dimensional system, simplifications are usually applied to the BS of anechoic and especially of reverberant situations. A reduction of complexity is possible by considering only virtual sources in the horizontal plane (cf. definition 2). Further, BS is often limited to BIRPs measured at a constant source distance, to one sound source radiation pattern, and to one source orientation (e. g. Mackensen et al. 1999). For the BS of reverberant environments, the number of head and source positions is typically also limited.

A further reduction of complexity is possible introducing unrealistic situations. If an exemplary system limited to one distance in the horizontal plane is considered, a typical BS of an unrealistic reverberant environment is implemented using only BIRPs representing a single source position and orientation combined with a single head position, but different head orientations. The BIRP update in the playback situation is hence done relative to the center of the listener's head. While this procedure would be correct for the BS of an anechoic situation, head-related BS using BIRPs measured in a reverberant environment is unrealistic. However, the implementation effort is reduced compared to the realistic room-related BS of the same reference scene. A variety of other unrealistic configurations for head and room-related BS is possible but not discussed in detail here. In general, unrealistic setups can provide good localization results and may, due to the achievable controllability, be helpful for perceptually addressing the performance of BS systems, as shown in the remainder of this section.

Grid Resolution Requirements Since spatial continuous systems are not possible with block-wise audio signal processing, it seems reasonable to ask in Zwicker and Feldtkeller's manner for the grid resolution perceptually not distinguishable from the reference scene, according to definition 19 referred to as transparent grid resolution (cf. section 1.1 and Zwicker and Feldtkeller 1967). A frequently used method of addressing auditory resolution psychoacoustically is the concept of the just noticeable sound changes (JNSCs), also referred to as minimum audible sound changes (cf. section 2.6 and Fastl and Zwicker 2007, p. 175). By definition, a transparent BS system provides the reference scene JNSCs. This constraint is not necessarily sufficient but required for transparent BS. Regarding the grid resolution, transparent BS must provide the reference scene JNSCs for source and listener movements. Consequently, a grid resolution at least providing the reference scene JNSCs is required. If at that resolution the switching between BIRPs is audible, a higher resolution is necessary (definition 29).

Definition 29 (*Grid Resolution for Transparent Binaural Synthesis*)

> *The grid resolution required for transparent binaural synthesis provides the reference scene just noticeable sound changes and artifact-free impulse response updates.*

Based on definition 29, inaudibility of artifacts or sound color changes due to BIRP changes is not considered sufficient for transparent BS in this work, in contrast to earlier studies on the grid or directional resolution of BS (e. g. Minnaar et al. 2005, Hoffmann and Møller 2008). Regarding the BS of anechoic environments, distinguishing head-related angular and distance resolution appears reasonable. For reverberant environments, a further distinction is necessary between the head-related angular resolution for a constant listener position and different angles between the listener's head and the source with regard to a fixed world coordinate system. The resolution required for angles between listener and source is given for static sources by the minimum audible angle (MAA) and for moving sources by the minimum audible movement angle (MAMA) according to section 2.6. Since the MAMA always exceeds the MAA (Chandler and Grantham 1992), the static situation represents the worst case. That being said, the head-related angular and distance resolutions remain to be addressed. In general, the grid resolution required for transparent BS depends on the specific system and the simulated environment.

In the following, methods of addressing the resolution requirements are proposed and verified. Further, interpolation methods for increasing the BIRP resolution are discussed, assuming the head tracking resolution and accuracy to be higher than the spatial grid resolution, which is typically fulfilled by the current technology[3] (Polhemus 2005).

Head-Related Angular Resolution Most earlier studies on the directional resolution required for transparent BS addressed the resolution providing the inaudibility of differences to a reference system. Minnaar et al. (2005) report for a specific static BS system, implemented based on free-field blocked auditory canal AH[4] recording and equalization,

[3] Polhemus 3 Space FasTrack, accuracy (RMS): 0.08 cm/0.15°, resolution: 0.05 cm/0.025°

[4] Department of Acoustics, Aalborg University, Denmark (VALDEMAR)

that the differences between resolutions of 2° and 4° are inaudible in horizontal and vertical direction. With a comparable BS system and the same AH, Hoffmann and Møller (2005a, 2006a, 2008) found a switching detection limit of about 2.8° for BIRPs with artificially removed interaural time differences (ITDs). Regarding ITD changes, Hall (1964) reports free-field detection thresholds for impulsive sounds of about 20 μs, also confirmed by Hoffmann and Møller (2005b, 2006b). This interaural delay corresponds to angular source position changes in the range of 2° (cf. Mills 1958). Lindau et al. (2008) concluded, based on dynamic BS experiments implemented with AH[5] blocked auditory canal recording in reverberant environment with 1° horizontal and vertical resolution, an angular grid resolution of 2° to enable a realistic synthesis of typical application scenarios.

Summarizing, earlier studies suggest an angular resolution in the range of 2° to be sufficient for the artifact-free BS of free-field as well as reverberant environments. Regarding source direction, the JNSCs are referred to as MAAs, with minimum values below 1°. In this work, the angular resolution requirement for not only artifact-free but also transparent BS is by definition 29 given as the resolution resulting in the MAAs of the *specific reference scene*. For determining the required resolution, the following procedure is proposed: The MAAs are measured according to section 2.6 in the reference scene and in BS situations with different resolutions. The required resolution is hence given by the maximum resolution resulting in the reference scene MAA (cf. Völk et al. 2012b).

The proposed procedure is verified by addressing the horizontal resolution required for the unrealistic room-related BS of a LS[6] in a reverberant listening environment. The system[7] is set up dynamically in the horizontal plane based on blocked auditory canal entrance artificial head recording as given by equation 4.53, using a head fixed with regard to the torso[8], and on average magnitude equalization according to equation 5.23. The resulting MAAs for different horizontal angular resolutions are shown by figure 5.1.

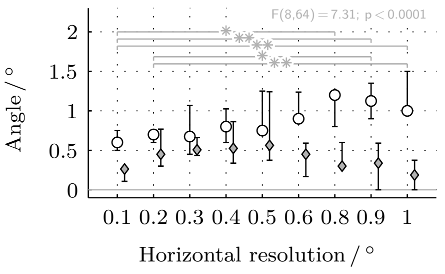

Figure 5.1: Quartiles of individual medians (open circles) and quartiles of individual interquartile ranges (filled rhombs) of minimum audible angles for room-related binaural synthesis in the horizontal plane with different angular resolutions (artificial head binaural impulse response set recorded in a reverberant laboratory environment, average headphone equalization, nine normal hearing subjects).

The resulting angles (open circles) indicate about 0.5° resolution to be sufficient for the MAA not to decrease further with the grid resolution and to reach the average MAA of about 0.7° reported for broadband stimuli played back by real sources or wave field synthesis in section 2.6. This interpretation is additionally supported by one factorial

[5] Department of Audio Communication, Technische Universität Berlin, Germany (FABIAN)

[6] Klein + Hummel Studio Monitor Loudspeaker O 98 at 2 m distance

[7] Sennheiser HD 800 headphones, $M_\mathrm{s} = 2$, $M_\mathrm{spo} = M_\mathrm{hpo} = M_\mathrm{so} = 1$, $M_\mathrm{ho} = 360, \ldots, 3600$

[8] Neumann KU 80

analysis of variance, indicating a highly significant effect of the horizontal angular resolution. A post-hoc comparison reveals significant differences only between the resolutions below 0.3° and above 0.6°, suggesting a transition region for the required resolution. Based on these results, 0.5° is proposed, while a strict quality criterion might require 0.3° to be selected as the horizontal angular resolution for the evaluated setup.

Figure 5.1 further shows that the intra-individual variability values (filled rhombs) depend on the grid resolution. This result is plausible based on the following hypothesis: At grid resolution values clearly exceeding the actual average intra-individual variability, small average intra-individual variability results are expected because the subjects are able to reliably discriminate two adjacent angular steps. As the grid resolution decays, a maximum of the intra-individual variability results is expected at the resolution representing the actual average intra-individual variability since the subjects are just not able to discriminate two adjacent angular steps. For resolutions clearly below the actual average intra-individual variability, the intra-individual variability results are expected to decay as the measurement gains accuracy. Consequently, the intra-individual variability characteristics shown by figure 5.1 support the discussion of the results given in the previous paragraph, especially the transition region of the required resolution between about 0.3° and 0.6°.

Because of the unrealistic situation simulated and the AH recording employed, the results given by figure 5.1 are not representative for arbitrary situations and BS with human head recording. However, using the setup and procedure introduced, the grid resolution required for a specific BS system can be determined. The similarity of the resulting angle to the MAA for broadband noise of about 0.7° indicates the correct order of magnitude and therefore the applicability of the procedure (cf. section 2.6). Considering individual BS and realistic reference scenes, it is important to take into account the procedural difficulties associated with high resolution measurements, especially on human subjects. For low resolutions, head movements during the measurements may exceed the grid resolution and thus corrupt the results (Christensen et al. 1999).

Head-Related Distance Resolution Otani et al. (2009) found the spectral shape of *free-field* BIRPs to depend on listener-source distances below about 2 m, in line with the data reported by Duda and Martens (1998). This distance dependence has been shown to be relevant for auditory localization (Brungart and Rabinowitz 1999, Brungart et al. 1999, Brungart 1999). At listener-source distances between 2 m and about 15 m, the free-field BIRPs for a constant direction show practically identical spectral shapes, differing only in the overall gain (Blauert and Braasch 2007, p. 14). At larger listener-source distances, the frequency dependent absorption of air becomes relevant (Evans et al. 1972, Zuckerwar and Meredith 1985). It is possible to incorporate the air absorption in BS using the frequency and distance dependent absorption coefficient formulated by Bass et al. (1990, 1995). In summary, continuous distance variations in BS of free-field situations can be implemented approximately correct for listener-source distances of more than about 2 m, while a discrete measurement grid is necessary for smaller distances.

In *reverberant environments*, geometry dependent reflection patterns occur in addition to the free-field effects. As a consequence, the BIRPs depend on the listener-source distance, theoretically requiring a discrete measurement grid at all distances.

To the author's best knowledge, no earlier studies exist discussing the JNSCs or the audibility of BIRP switching artifacts regarding the listener-source distance in BS. For that reason, the two-step procedure introduced in the previous section for determining the necessary angular BS resolution is extended to the distance resolution: In a first step, the grid resolution providing the reference scene JNSCs regarding listener-source distance is determined. If necessary, the grid resolution is decreased further in a second step until no artifacts are audible on switching between BIRPs for different distances. Since the listener-source distance related JNSCs are the distance dependent minimum audible distances (MADs) discussed in section 2.6, the grid resolution resulting in the reference scene MADs depends on the listener-source distance and the reference scene. For that reason, the evaluation procedure proposed must be executed for the specific reference scene to be synthesized.

Interpolation If the effort of recording or synthesizing all BIRPs required for the selected combination of BS approach and reference scene is not justified or possible, interpolation procedures are frequently used to increase the BIRP grid resolution. A typical procedure is based on representing the BIRPs by their minimum-phase components and a pure delay and on interpolating both components separately (Kistler and Wightman 1992, Jot et al. 1995, Kulkarni et al. 1999, Savioja 2000, Wenzel et al. 2000). Plogsties et al. (2000) showed the BIRP partitioning in minimum-phase and delay contributions not to be distinguishable from the original if the low-frequency group delay is taken as the delay contribution, which is supported by Kulkarni et al. (1999). Another frequently employed interpolation procedure is to time align the BIRPs and to interpolate the initial delay and the time aligned contributions separately (e. g. Djelani et al. 2000).

Christensen et al. (1999) found the horizontal angular grid resolution of *free-field* AH BIRPs required as an interpolation basis to depend on the interpolation method, while observing a general tendency of larger interpolation errors on the contra-lateral side. Linearly interpolating the logarithmic magnitude values and phase angles of the BIRPs separately, Langendijk and Bronkhorst (2000) show the individual dynamic BS[9] of an LS in the free field not to be distinguishable from the corresponding reference scene for interpolations based on angular grid resolutions up to 6°. With larger supporting angle densities, at first coloration becomes audible, and if the supporting angle density exceeds about 20°, the localization is affected.

Not all temporal sections of BIRPs representing *reverberant conditions* may be of equal perceptual relevance (Jensen and Welti 2003). Therefore, a mathematically correct interpolation is not necessarily required for all directions, distances, or BIRP sections. Restricting the interpolation to the perceptually most relevant components can save computational power in real-time implementations. The processing strategy applied to the components considered less relevant may influence the audibility thresholds of the procedure. For example, Meesawat and Hammershøi (2003) have shown keeping the reverberation in the BIRPs after some 100 ms constant for all directions to be inaudible.

[9] Sony MDR E-575 headphones mounted approximately 1 cm outside the entrance to the auditory canal, human head probe microphone recording and headphone transfer functions measurement

5.1.3 Nonindividual Recording

Correct BS requires, following section 4.3, individual recording. To the author's best knowledge, all studies on the influence of nonindividual BIRPs on the localization in BS agree that individual recording allows for the highest synthesis quality, regardless of the recording method (Wenzel et al. 1991, Bronkhorst 1995, Møller et al. 1996c). This fact appears plausible taking into account the inter-individual variation of the physical BIRP characteristics (Mehrgardt 1975, Middlebrooks et al. 1989, Carlile and Pralong 1994, Wightman and Kistler 1996, Hammershøi and Møller 1996b).

Nonindividual Human Head Recording Møller et al. (1996c) studied the ability to correctly identify one of nineteen LSs as the source of a female speech excerpt. The LSs were distributed in a reverberant environment visibly at different angles in azimuth and elevation around the subjects. Møller et al. included three playback methods in the study: the real sources, the corresponding individual binaural recordings, and the corresponding nonindividual binaural recordings of a randomly selected subject[10]. With individual recording, the amount of errors remained comparable to the real source condition, while with nonindividual recording the numbers of elevation errors and front-back confusions increased. Thereby, frontally synthesized sources tended to appear in the back, and sources intended in the horizontal plane were frequently perceived too high. These results were confirmed with static BS[11] by Wenzel et al. (1993) and Seeber (2003, pp. 52–61), who also found an increased number of in-the-head localizations[12]. Møller et al. (1996c) further showed the number of distance errors to increase with nonindividual instead of individual binaural recording, supported with static BS[13] by Kim and Choi (2005).

The average number of localization errors with nonindividual recording can be reduced by selecting the BIRPs, which holds true for BS with blocked auditory canal (Møller et al. 1996a,b) and with probe microphone recording (Wenzel et al. 1993). However, Seeber (2002a) found a highly significant increase of the frontal horizontal plane localization variability for static BS[14] with selected nonindividual versus individual free-field recording. Seeber and Fastl (2001, 2003) proposed a procedure for perceptively selecting the BIRP suited best regarding frontal horizontal plane localization accuracy. The selection criterion of Seeber and Fastl was extended by Iwaya (2006) to three-dimensional scenarios. Schönstein and Katz (2010) proposed a BIRP selection procedure based on morphologic parameters, which has been shown to predict perceptual ranking results significantly better than random choice. An alternative to the BIRP selection is their customization (Martens 2003, Xu et al. 2007). Examples of customization approaches are the scaling in frequency (Middlebrooks 1999b,a, Middlebrooks et al. 2000), perceptually motivated primary component analysis and synthesis (Hwang and Park 2007, Hwang et al. 2010), and perceptually motivated spectral smoothing and equalization (Silzle 2002a,b).

[10] Beyerdynamic DT 990 headphones, blocked auditory canal recording, individual magnitude equalization

[11] Sennheiser HD 430 headphones, probe microphone recording, nonindividual magnitude equalization

[12] Stax SR λ pro headphones, blocked auditory canal recording, no equalization

[13] Sennheiser HD 600 headphones, probe microphone recording, individual equalization

[14] Stax SR λ pro headphones, blocked auditory canal recording, no equalization

Artificial Head Recording Burkhard and Sachs introduced in 1975 a manikin for acoustic research[15], designed to match average anthropometric data regarding head, torso, auricle, auditory canal, and eardrum. However, the manikin's surface and skin impedances do not emulate human data (Burkhard and Sachs 1975). Assumption 8 requires the AH for the BIRP recording in BS to represent the listener's head regarding HP and LS playback. This assumption is in general still not fulfilled for current AHs (Daniel et al. 2007).

Møller et al. (1997, 1999) studied the errors of localizing LSs distributed visibly at nineteen different positions in azimuth and elevation around the subjects in a reverberant environment using a female speech stimulus. Playback conditions included were the real sources and the corresponding static BS[16], implemented based on recording with different AH models[17]. BS with standard AH and with blocked auditory canal entrance AH recording resulted in an increased number of localization errors compared to the real sources, with a raise of median plane errors by more than 100%. On average, the numbers of errors for both recording methods tend to be worse than with nonindividual recording using a randomly selected human subject (cf. Møller et al. 1996c). In a repetition study with different playback equipment[18] and including newer AHs[19], Minnaar et al. (2001, 2004) in general confirmed the results of Møller et al. (1999). With BS, the best additionally included AHs provided a number of median plane errors between the numbers achieved with arbitrary and selected nonindividual recording, which rebuts assumption 8 for current AHs. In the course of the experiments, a torso simulation, providing shoulder reflections, was identified important (Minnaar et al. 2001). Surveys comparing current AH dimensions with anthropometric data indicate a geometric mismatch, which is likely to contribute to the deviations in localization performance of human and artificial heads (Daniel et al. 2007, Genuit and Fiebig 2007).

Common Aspects Silzle (2002a,b) found artificial versus human head BIRPs to sound less natural in BS with various HP models, AHs, and recording procedures, even when including a perceptually motivated equalization. In contrast, for comparisons of AH and individual recording in static and dynamic BS[20] regarding localization, externalization, and realism, Begault et al. (2001) reported no significant main effect of the recording method. However, their results may be deteriorated by the suboptimal equalization. Usher and Martens (2007) found nonindividual BIRPs in BS[21] rated more natural than individual BIRPs. This at first glance unexpected result may be understood taking into

[15] KEMAR artificial head

[16] Beyerdynamic DT 990 headphones, standard and blocked auditory canal artificial head recording, individual and nonindividual equalization

[17] KEMAR artificial head with different pinnae, Neumann KU 80i and 81i, Head Acoustics HMS I and II, Brüel & Kjær 4128 and 5930, Inst. of Biomed. Eng. at Univ. of Toronto

[18] Sennheiser HE 60 headphones, standard artificial head recording, nonindividual equalization

[19] KEMAR artificial head, Neumann KU 100, Brüel & Kjær 4100, Head Acoustics HMS II, Cortex Electronic MK 1, Institute of Technical Acoustics, Aachen University, Germany (ITA), Department of Acoustics, Aalborg University, Denmark (VALDEMAR)

[20] Sennheiser HD 430 diffuse-field equalized headphones with AH and individual blocked auditory canal recording, no further equalization

[21] iPod 12-2006 headphones, blocked auditory canal recording (reciprocity technique), no equalization

account that the most natural situation not necessarily equals the most authentic situation (cf. section 3.1.1). Therefore, naturalness is not suited as a quality criterion for BS aiming at reproducing a reference scene. Having said this, the need for a perceptual BS quality criterion more selective than the accuracy of the absolute localization, as requested by Martens et al. (2010), is supported by results of this thesis. The criterion proposed here is the frequency independence of the loudness transfer function (LTF) measured for the BS system under consideration with regard to the reference scene (cf. section 2.5). A study addressing the impact of individual recording and equalization on the LTF of the BS of an exemplary LS reference scene is discussed in section 5.4. The findings indicate degraded loudness transfer for nonindividual versus individual recording and equalization.

The results of earlier perceptual BS evaluations regarding individual versus nonindividual recording can be summarized in that localization accuracy and loudness transfer sequentially decay when going from real sources to individual BS, to nonindividual BS with selected human head BIRPs, and to nonindividual BS with arbitrary human or artificial head BIRPs. Consequently, current research aims at fast BIRP measurement, which can reduce the effort of implementing individual BS (e. g. Enzner et al. 2011).

5.1.4 Probe Microphone Ear Signal Measurement

According to assumption 4, the sound pressure detected by a probe microphone tube tip close to the eardrum represents the sound pressure actually evaluated by the hearing system, the so-called ear signal (cf. definition 22). Stinson (1985) showed, by simulations and upscale model measurements, that the sound pressure varies over the eardrum surface by more than 20 dB at frequencies above about 6 kHz (confirmed by Stinson and Khanna 1989, Stinson and Lawton 1989, Neely and Gorga 1998, Stinson and Daigle 2005, and Hudde and Schmidt 2009). At frequencies below some 6 kHz, probe tube tip distances up to about 5 mm, measured along the canal axes outward from the orthogonal projection of the top of the eardrum, allow for results representing the eardrum pressure with an accuracy of ±3 dB (Chan and Geisler 1990, Hellstrom and Axelsson 1993). However, a probe tube inserted in the auditory canal influences the sound field in the canal by its presence, shows frequency dependent transfer characteristics, and provides a reduced dynamic range compared to measurement microphones (Hellstrom and Axelsson 1993) and miniature microphones (Møller et al. 1995a). The complex high frequency structure of the canal sound field is shown further by finite element simulations of Schmidt and Hudde (2009), indicating the results of eardrum impedance estimates based on measurements at remote points in the canal to be questionable at frequencies above 3 kHz, due to an inaccurate transformation to the eardrum. If the auditory canal is occluded, for example by an HP, standing waves may introduce further errors (Richmond et al. 2011).

Summarizing, thoroughly conducted probe microphone ear signal approximations may represent the actual eardrum pressure in the frequency range below some 3 kHz, but at higher frequencies deviations up to 20 dB occur, primarily due to two effects: the probe tube tip cannot be positioned directly at the eardrum, and the sound pressure varies across the eardrum surface and is therefore not captured in an identical manner by the eardrum and the typically smaller probe tube tip. As a consequence, assumption 4 never holds true for the whole audible frequency range, but may be fulfilled using a thorough procedure at frequencies below about 3 kHz. However, BS with probe microphone recording represents the initial approach to BS (Wightman and Kistler 1989a), has been in use constantly since then (Griesinger 1990, Langendijk and Bronkhorst 2000), and is still the preferred procedure of some researchers (e. g. Griesinger 2009).

5.1.5 Transfer Function Smoothing

Typically, the magnitude spectra of the recording situation transfer functions (TFs) defined by equation 4.12 show strong variation over frequency. An exemplary recording situation TF measured in a reverberant reference scene is depicted by figure 5.2.

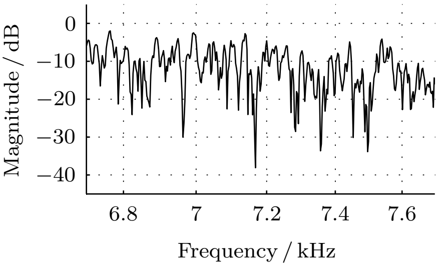

Figure 5.2: Excerpt from the magnitude spectrum of a typical blocked auditory canal artificial head recording situation transfer function before smoothing. The prominent peaks and dips form the motivation for spectral smoothing.

Pronounced spectral variation may induce technical problems when implementing BS since large linear amplification is required to play back signals convolved with the corresponding IRs. The amplification required can be reduced by spectral smoothing. Dips and especially zeros in the magnitude spectra may also prevent or complicate spectral inversion, as for example required for the equalization of BS (cf. assumption 3). Spectral smoothing reduces dips, limits the number of zeros, and has proven useful to cope with spectral inversion problems in general (cf. section 5.2.3).

In the context of audio systems, the goal of TF smoothing is to reduce spectral fluctuation without diminishing the amount of relevant auditory information. Aiming at this goal, the system-theoretic foundation of a smoothing procedure referred to as auditory-adapted exponential transfer function smoothing (AAS) is derived in the following. AAS is adapted to the spectral properties of the human hearing system by implementing frequency dependent smoothing bandwidths (cf. Völk et al. 2011a). Possibly occurring time domain effects are identified and discussed along with two procedures to cope with the resulting shortcomings. The section is concluded by listening experiments determining audibility thresholds of AAS and the perceptual manifestation of audible spectral smoothing.

System-Theoretic Basics of Spectral Smoothing in General Spectral smoothing is possible by convolving the spectrum of the signal to be smoothed with a spectral kernel (Hatziantoniou and Mourjopoulos 2000a,b). This procedure is referred to as linear smoothing, in contrast to cepstral or wavelet based methods (Hacihabiboglu et al. 2002). Considering a continuous IR on a continuous and infinite time scale, convolution in the frequency domain equals, according to the convolution theorem (Oppenheim et al. 1998), multiplication of the signal with a window function in the time domain. Accordingly, the smoothing can be specified either in the frequency domain by a spectral kernel or in the time domain by a window function. Since acoustic systems in general feature complex-valued TFs, the frequency domain convolution with a spectral kernel affects the real and imaginary parts of the TF. For spectral smoothing, the kernel is intended to meet the following target requirements as closely as possible:

- The spectral kernel should feature a real-valued time domain representation decaying monotonously within the relevant temporal window (cf. Mummert 1997).
- The spectral kernel should exhibit spectral symmetry with a maximum at 0 Hz, so that no energy is systematically shifted along the frequency scale.

In contrast to the theoretical continuous case regarded so far, recorded IRs are typically available windowed in the time domain and digitized, that is discrete in value and time. In the digital case, the temporal and spectral periodicity of the discrete Fourier transform (DFT) pair has to be considered in the smoothing process. Thus, circular convolution, denoted by $\circledast$, has to be used to apply the kernel in the frequency domain, represented by element-wise multiplication of the corresponding time series. The recorded IR

$$h\,[n] = \sum_{i=-\infty}^{\infty} h\,[i]\,\delta[n-i], \quad n = 0, \ldots, N-1 \tag{5.5}$$

is given as a sequence of discrete sample values. The associated frequency domain representation $H\,[k]$, with $k = 0, \ldots, N-1$, is computed by DFT (Oppenheim et al. 1999, pp. 559–564). In the following, $H\,[k]$ represents the TF to be smoothed.

Previous Approaches to Spectral Smoothing The most direct approach to spectral smoothing is averaging the weights of the spectral TF bins in a symmetric region around each bin, in the time domain represented by multiplying the IR with a window function (Hatziantoniou and Mourjopoulos 2000b). This approach reduces the amount of auditory useful information in an acoustic IR, for example by degrading the late reverberation in a room IR. However, the procedure is discussed since it represents the basic scenario for more elaborate linear smoothing approaches. In the frequency domain, the situation can be described as the convolution of $H\,[k]$ with the symmetric rectangular spectral kernel

$$W\,[k] = \begin{cases} \frac{1}{2m+1} & \text{for} \quad 0 \le k \le m \;\; \text{and} \;\; N-m \le k \le N-1, \quad \text{with } 2m < N, \\ 0 & \text{otherwise.} \end{cases} \tag{5.6}$$

The integer m determines the kernel bandwidth and therefore controls the smoothing resolution. The corresponding time domain window function is given by

$$w\,[n] = \frac{1}{N(2m+1)} \frac{\sin\left(\pi n(2m+1)/N\right)}{\sin(\pi n/N)}\,, \quad \text{with } n = 0, \ldots, N-1 \text{ and } 2m < N. \tag{5.7}$$

All earlier approaches to linear smoothing known to the author are based on linear-phase spectral kernels, typically specified in the frequency domain (an overview of smoothing methods is presented by Hatziantoniou and Mourjopoulos 2000b). Using a real-valued spectral kernel, the real and imaginary parts of the TF are averaged independently, and the corresponding time domain window is symmetric to its midpoint. As a result, the mid part of the IR is attenuated by the smoothing, while the initial and final sections are only slightly modified. Possibly occurring artifacts due to the increasing amplitude of the window function at the end of the IR can be reduced by temporal zero padding or additional pre-smoothing (Hatziantoniou and Mourjopoulos 2000b). Non-monotonic temporal amplitude decay as occurring for the rectangular spectral kernel introduces further undesired artifacts, making the choice of an appropriate window function an essential step in the design process.

For auditory-adapted smoothing, a frequency dependent spectral resolution is desirable, implemented by kernel bandwidths depending on the analysis frequency bin κ. The frequency dependent resolution is motivated by the fact that the hearing system subdivides sound intensity spectrally non-uniformly, as discussed in section 2.3. Typically, fractional-octave functions are used for smoothing (Hatziantoniou and Mourjopoulos 2000b), while aiming at a better auditory adaption the critical-band concept, reviewed and reformulated in section 2.3, is preferred here (cf. Zwicker 1961b). Since it is desired not to systematically shift energy in the frequency domain, all frequency dependent weights $W_\kappa\,[k]$ must be centered symmetrically on κ. This is implemented replacing the constant factor m of equation 5.6 by the frequency dependent factor m_κ. A TF can be written using a series of non-overlapping, directly neighboring spectral windows $S_\kappa\,[k]$ by

$$H\,[k] = \sum_\kappa H\,[k]\,S_\kappa\,[k]\,. \tag{5.8}$$

Each window has unity weights for all bins within m_κ and equals zero otherwise. The most extreme situation consists of N windows $S_\kappa\,[k]$, each representing a one spectral line wide band-pass filter with the IR $s_\kappa\,[n]$. The spectrally smoothed TF

$$H_{\mathrm{sm}}\,[k] = \sum_\kappa H\,[k]\,S_\kappa\,[k] \circledast W_\kappa\,[k] \tag{5.9}$$

is represented in the time domain by the IR

$$h_{\mathrm{sm}}\,[n] = \sum_\kappa \Big(h\,[n] \circledast s_\kappa\,[n]\Big) w_\kappa\,[n]\,, \tag{5.10}$$

computed by band-pass filtering $h\,[n]$ with all band-pass filters $s_\kappa\,[n]$ and subsequent summation of the time domain results.

Previous approaches to linear spectral smoothing known to the author can be assigned to the following four categories:

- *Complex Smoothing:* Convolution of the TF with a linear, real-valued spectral kernel (Hatziantoniou and Mourjopoulos 2000b).

- *Power Smoothing:* Convolution of the magnitude spectrum with a linear, real-valued kernel, that is smoothing the squared magnitude $|H[k]|^2$ and combining the result with the original phase $\arg(H[k])$ (Hatziantoniou and Mourjopoulos 2000b).

- *Equivalent Complex Smoothing:* Combination of the power smoothed magnitude with the complex smoothed phase (Hatziantoniou and Mourjopoulos 2000b).

- *Magnitude and Phase Smoothing:* Averaging magnitude and phase separately (Panzer and Ferekidis 2004).

Auditory-Adapted Exponential Transfer Function Smoothing (AAS) Using complex-valued spectral kernels, the real and imaginary parts of the TF are no longer smoothed independently. Similar to the Fourier-t transform (FTT) defined by Terhardt (1985), exponentially decaying temporal windows are applied, represented discretely by

$$w_{\mathrm{e}_\kappa}[n] = \mathrm{e}^{-\frac{n a_\kappa}{f_\mathrm{s}}} \varepsilon[n] \quad \circ\!\!-\!\!\!-\!\!\bullet \quad W_{\mathrm{e}_\kappa}[k] = \frac{1 - \mathrm{e}^{-\frac{N a_\kappa}{f_\mathrm{s}}}}{1 - \mathrm{e}^{-\mathrm{j}\frac{2\pi k}{N} - \frac{a_\kappa}{f_\mathrm{s}}}}. \tag{5.11}$$

Thereby, $a_\kappa > 0$ denotes time constants dependent on the analysis frequency bin κ, which allow for adapting the spectral smoothing resolution to the critical bandwidth (CBW). Further, $\varepsilon[n]$ denotes the unit step sequence according to Oppenheim et al. (1998, equation 1.64), and f_s represents the sample rate. The derivation of the discrete time representation is given in appendix D.1. The smoothing windows are selected so that no weighting occurs for $n = 0$, that is $w_{e_\kappa}[0] = 1 \; \forall\, \kappa$ holds true. In contrast to the FTT, the windows decay with increasing time. In the frequency domain, AAS equals the convolution with the frequency dependent complex-valued smoothing kernels $W_{\mathrm{e}_\kappa}[k]$, according to Terhardt (1985) featuring the 3 dB-bandwidth

$$\Delta f_\kappa = a_\kappa / \pi. \tag{5.12}$$

Consequently, the degree of AAS is controlled by the frequency independent parameter c_sm, modifying the spectral kernel bandwidth by the window function time constant

$$a_\kappa = c_\mathrm{sm} \cdot \Delta f_\mathrm{Gv}(\kappa f_\mathrm{s}/N), \tag{5.13}$$

set to a multiple of the CBW $\Delta f_\mathrm{Gv}(f)$ according to equation 2.11. To preserve a high resolution of the smoothing frequency dependence, $S_\kappa[k]$ is selected spectrally as narrow as possible, that is one spectral line wide, represented by the band-pass filter series

$$S_{\mathrm{e}_\kappa}[k] = \delta[k-\kappa], \quad \kappa = 0, \ldots, N-1. \tag{5.14}$$

Equation 5.9 can be written using equation 5.14 to describe AAS for $N > 1$ by

$$H_\mathrm{sm}[k] = \sum_{\kappa=0}^{N-1} H[\kappa]\, \delta[k-\kappa] \circledast W_{\mathrm{e}_\kappa}[k]. \tag{5.15}$$

Combined with the DFT definition (Oppenheim et al. 1999, equations 8.67 and 8.68), the synthesis equation for the IR smoothed by AAS is given by

$$h_{\mathrm{sm}}[n] = \frac{1}{N}\sum_{k=0}^{N-1} \mathrm{e}^{\mathrm{j}\frac{2\pi nk}{N}} H_{\mathrm{sm}}[k] = \frac{1}{N}\sum_{k=0}^{N-1} \mathrm{e}^{\mathrm{j}\frac{2\pi nk}{N}} \sum_{\kappa=0}^{N-1} H[\kappa]\Big(\delta[k-\kappa] \circledast W_{\mathrm{e}_\kappa}[k]\Big). \tag{5.16}$$

Using the DFT re-synthesis definition (Oppenheim et al. 1999, equation 8.68) and the spectral shifting property of the DFT, equation 5.16 can be written by

$$\begin{aligned} h_{\mathrm{sm}}[n] &= \sum_{\kappa=0}^{N-1} H[\kappa]\frac{1}{N}\sum_{k=0}^{N-1} \mathrm{e}^{\mathrm{j}\frac{2\pi nk}{N}}\Big(\delta[k-\kappa] \circledast W_{\mathrm{e}_\kappa}[k]\Big) \\ &= \sum_{\kappa=0}^{N-1} H[\kappa]\left(\frac{1}{N}\mathrm{e}^{\mathrm{j}\frac{2\pi\kappa n}{N}} w_{\mathrm{e}_\kappa}[n]\right) = \frac{1}{N}\sum_{\kappa=0}^{N-1} \mathrm{e}^{\mathrm{j}\frac{2\pi\kappa n}{N}} H[\kappa]\, w_{\mathrm{e}_\kappa}[n]. \end{aligned} \tag{5.17}$$

Accordingly, AAS can be understood as temporally weighted DFT-re-synthesis. Rewriting the TF $H[\kappa]$ in polar form, equation 5.17 can be simplified to

$$h_{\mathrm{sm}}[n] = \frac{1}{N}\sum_{\kappa=0}^{N-1} \big|H[\kappa]\big|\,\mathrm{e}^{\mathrm{j}\varphi[n,\kappa]} w_{\mathrm{e}_\kappa}[n], \quad \varphi[n,\kappa] = \frac{2\pi\kappa n}{N} + \arg\big(H[\kappa]\big). \tag{5.18}$$

As $H[k]$ is computed by DFT from a real-valued time function, it is Hermitian and

$$H[k] = H^*[N-k] \tag{5.19}$$

holds true (Oppenheim et al. 1999, p. 576). Assuming sequences of even sample numbers N (odd numbered sequences may without loss of generality be zero-padded, Kammeyer and Kroschel 2006, p. 233), equation 5.18 is further simplified to

$$\begin{aligned} h_{\mathrm{sm}}[n] = {} & \frac{H[0]\, w_{\mathrm{e}_0}[n]}{N} + \sum_{\kappa=1}^{\frac{N}{2}-1} \frac{\big|H[\kappa]\big| w_{\mathrm{e}_\kappa}[n]}{N}\Big(\mathrm{e}^{\mathrm{j}\varphi[n,\kappa]} + \mathrm{e}^{-\mathrm{j}\varphi[n,\kappa]}\Big) + \\ & + \frac{H\left[\frac{N}{2}\right] w_{\mathrm{e}_{\frac{N}{2}}}[n]}{N}\,\mathrm{e}^{\mathrm{j}n\pi}. \end{aligned} \tag{5.20}$$

With $\cos\varphi[n,\kappa] = 0.5\,\mathrm{e}^{\mathrm{j}\varphi[n,\kappa]} + 0.5\,\mathrm{e}^{-\mathrm{j}\varphi[n,\kappa]}$ (Abramowitz and Stegun 1972, equation 4.3.2) and re-substitution, the synthesis equation of the spectrally smoothed IR is given by

$$\begin{aligned} h_{\mathrm{sm}}[n] = \frac{1}{N}\Bigg[& H[0]\, w_{\mathrm{e}_0}[n] + 2\sum_{\kappa=1}^{\frac{N}{2}-1} \big|H[\kappa]\big| w_{\mathrm{e}_\kappa}[n]\cos\left(\frac{2\pi\kappa n}{N} + \arg\big(H[\kappa]\big)\right) + \\ & + H\left[\tfrac{N}{2}\right] w_{\mathrm{e}_{\frac{N}{2}}}[n]\cos\big(n\pi\big)\Bigg]. \end{aligned} \tag{5.21}$$

Time Domain Aspects A temporal offset between the IR to be spectrally smoothed and the AAS window function influences the smoothing result since spectral smoothing equals a multiplication of the IR to be smoothed with a decaying temporal window function, independently of the smoothing method. Therefore, different initial delays in an otherwise unchanged IR lead to different smoothing results. For example, the frequency dependent smoothing of an IR with initial delay results in undesired damping at frequencies where the window function decays considerably before the first impulse occurs. Ideally, the attenuation should be minimal for the initial impulse, which is only provided if the window function is aligned with the IR. Two approaches of AAS for IRs with an initial delay are proposed: circular temporal shifting and minimum-phase/all-pass decomposition.

Circular temporal shifting means circularly rotating the IR prior to the smoothing so that the first impulse is located at the beginning of the IR and the initial delay is shifted to the end. As a consequence, the first impulse and the maximum of the window functions coincide and subsequent signal contributions are damped dependent on frequency. Afterwards, the result is circularly shifted back. A difficulty regarding this procedure is the identification of the first impulse within the IR. For the AAS applications discussed in the following, a heuristic procedure is used, detecting the first signal slope larger than a multiple of the average noise amplitude, selected dependent on the recording situation.

An approach referred to as *minimum-phase/all-pass decomposition* is discussed in principle, for being independent of the initial delay. Using the decomposition $H[k] = H_{\mathrm{mp}}[k] \cdot H_{\mathrm{ap}}[k]$, the minimum-phase contribution $H_{\mathrm{mp}}[k]$ and the all-pass contribution $H_{\mathrm{ap}}[k]$ are separated. Following Mehrgardt and Mellert (1977) only $H_{\mathrm{mp}}[k]$ is smoothed, resulting in $H_{\mathrm{mp,sm}}[k]$. The energy maximum of a minimum-phase system occurs with the least temporal delay of all causal systems with the same magnitude spectrum (Oppenheim et al. 1999, pp. 280–291). Consequently, all high energy signal contributions are shifted towards the beginning of the minimum-phase IR, while their original positions are encoded in the all-pass contribution. As a consequence, less unintended attenuation occurs when smoothing the minimum-phase part. After the smoothing, the original temporal distribution is reintroduced by the multiplication $H_{\mathrm{sm}}[k] = H_{\mathrm{mp,sm}}[k] \cdot H_{\mathrm{ap}}[k]$. This reconstruction may induce time domain artifacts. Since the artifacts typically occur in the late part of the IR, they can be attenuated by additional temporal windowing.

Perceptual Consequences of Spectral Smoothing Applied to BIRPs, spectral smoothing may become audible in BS for example by changing the sound color or affecting the localization. The audibility of AAS induced sound color changes is studied by applying AAS according to equation 5.21 to the finite impulse response (FIR) approximations of two IRs of an LS[22], one IR measured in an anechoic chamber with 250 Hz lower limiting frequency, the other in a reverberant laboratory room with 500 ms average reverberation time. The spectrally smoothed and the original FIRs were convolved with broadband uniform exciting noise (UEN) impulses[23] and examined in a listening experiment. UEN impulses were shown in a pre-test to allow for more critical sound color evaluations than

[22] Klein + Hummel Studio Monitor Loudspeaker O 98

[23] 700 ms impulse duration, 5 ms Gaussian gating, UEN according to Fastl and Zwicker (2007, p. 170)

speech or impulsive sounds. Using a three-alternative forced choice procedure (Hellbrück and Ellermeier 2004, p. 226) combined with a 2-down 1-up adaption rule (Levitt 1971), the just noticeable AAS smoothing bandwidth Δf_κ according to equation 5.12 is determined for diotic free-field equalized HP playback[24] at a free-field equivalent level of 70 dB SPL. Table 5.1 shows the resulting quartiles of the intra-individual medians of eleven normal hearing subjects as a fraction of the CBW $\Delta f_\mathrm{G} = \Delta f_{\mathrm{G}_\mathrm{Z}}\,(\kappa f_\mathrm{s}/N)$, computed according to equation 2.9, to be comparable to the discussion of Fastl and Zwicker (2007, pp. 194–200).

percentile	25%	50%	75%
anechoic chamber	$\Delta f_\mathrm{G}/30$	$\Delta f_\mathrm{G}/21$	$\Delta f_\mathrm{G}/19$
laboratory room	$\Delta f_\mathrm{G}/198$	$\Delta f_\mathrm{G}/177$	$\Delta f_\mathrm{G}/150$

Table 5.1: Inter-individual statistics of the bandwidth Δf_κ for just noticeable auditory-adapted exponential smoothing of a loudspeaker transfer function measured in two rooms with different acoustical conditions as a fraction of the critical bandwidth Δf_G.

For the anechoic environment, the results correspond closely to the just noticeable variation in frequency $2\Delta f = \Delta f_\mathrm{G}/27$ reported for pure tones by Fastl and Zwicker (2007, equation 7.7). In the reverberant condition, AAS is perceivable at about seven to eight times smaller bandwidths. This result confirms that the spectro-temporally effective TF of the LS-room system is processed by the auditory system differently than the primarily spectrally effective LS TF recorded in the anechoic chamber (cf. section 2.4).

Regarding the consequences of spectral smoothing on the auditory localization, Kulkarni and Colburn (1998) showed the hearing sensation positions in static BS[25] of free-field conditions, set up with individual blocked auditory canal entrance recording, to be robust against spectral smoothing. In Kulkarni and Colburn's study, the smoothing became, for increasing bandwidths, physically effective before affecting the auditory localization at which the hearing sensation elevation was most likely influenced. This result was confirmed with static BS by Breebaart et al. (2010)[26] and Xie and Zhang (2010)[27].

In the course of this work, AAS induced localization effects are evaluated using a broadband UEN pulse[28] presented by a virtual LS[29], binaurally synthesized in front of the subjects in the horizontal plane. The dynamic BS[30] was implemented with nonindividual blocked auditory canal entrance recording and average magnitude equalization, taking into account four BIRP conditions: original and smoothed by AAS with the three bandwidths Δf_{κ_1}, Δf_{κ_2}, and Δf_{κ_3} given by table 5.2. The smoothing bandwidths were selected with regard to the audibility of sound color changes: According to table 5.1, AAS with Δf_{κ_1} causes for free-field equalized diotic HP presentation no sound color

[24] Beyer DT 48 with passive free-field equalizer according to Fastl and Zwicker (2007, p. 7)

[25] Etymotic Research ER-2 tube-phones without foam-tips, individual equalization

[26] Beyerdynamic DT 990 headphones, nonindividual blocked auditory canal recording, no equalization

[27] Sennheiser 250 Linear II headphones, individual blocked auditory canal recording and equalization

[28] 700 ms impulse duration, 5 ms Gaussian gating, 300 ms pause

[29] Klein + Hummel Studio Monitor Loudspeaker O 200

[30] Stax λ pro NEW headphones, $M_\mathrm{s} = 1$, $M_\mathrm{spo} = M_\mathrm{hpo} = M_\mathrm{so} = 1$, $M_\mathrm{ho} = 360$

change with regard to the non-smoothed condition in both binaurally synthesized acoustic environments. In contrast, Δf_{κ_2} and Δf_{κ_3} result in audible sound color changes in the synthesized reverberant laboratory, and Δf_{κ_3} even in the synthesized anechoic chamber.

identifier	Ref	Δf_{κ_1}	Δf_{κ_2}	Δf_{κ_3}
bandwidth	no AAS	$\Delta f_{\mathrm{G}}/233$	$\Delta f_{\mathrm{G}}/117$	$\Delta f_{\mathrm{G}}/12$

Table 5.2: Auditory-adapted exponential transfer function smoothing (AAS) bandwidths employed to study the influence of AAS on the auditory localization, given as a fraction of the critical bandwidth Δf_{G} according to equation 2.9.

In the localization experiment, eleven normal hearing subjects were asked to verbally evaluate the following hearing sensation properties three times per AAS bandwidth condition: azimuth angle with regard to the median plane, height with regard to the horizontal plane, distance to the center of the head, and horizontal width.

The results of the experiment reveal no significant influence of the factor AAS bandwidth regarding azimuth [F(3,30) = 2.53; p = 0.0758] and height [F(3,30) = 0.16; p = 0.9226] perception, whereas hearing sensation distance [F(3,30) = 11.89; p < 0.0001] and width [F(3,30) = 8.44; p = 0.0003] are highly significantly influenced by AAS. A post-hoc comparison reveals the smoothing with Δf_{κ_3} to differ from the other conditions: by AAS with Δf_{κ_3}, distance ratings are reduced by about 50% and width judgments by some 30%.

Summarizing, AAS applied to BIRPs in BS can affect sound color and auditory localization. The sound color is more likely affected for the synthesis of reverberant versus anechoic conditions. If the degree of smoothing is increased, sound color deviations from the non-smoothed condition become audible before the localization is affected. Regarding the localization, the azimuthal localization appears most robust, while height, width, and distance judgments are more likely to be affected by spectral smoothing.

5.2 Aspects of the Equalization Procedure

Chapter 4 shows that the reference scene ear signals can be recreated approximately by BS if assumptions 5 and 6 are met, meaning the HPTFs are measured according to definition 26 individually at the positions of the recording microphones and the HP positions are identical for playback and HPTF measurement. If the recording is carried out with miniature microphones at the blocked auditory canal entrances, it is further necessary to employ appropriate HPs, as defined in section 5.3. In line with these results, Wightman and Kistler (2005, pp. 434–435) stressed the general importance of the BS equalization, especially when BS is employed for the audio playback in hearing research, since non-equalized HPs may introduce spectral cues that might erroneously be interpreted by the hearing system as localization cues.

Wightman and Kistler's argumentation may be generalized by requesting a clear distinction between physical stimuli and auditory perceptions (cf. section 3.2): Stimulation with physically defined ear signals by BS requires the system-theoretically correct equalization derived in chapter 4. Pointing out this fact appears particularly important considering the correct equalization is frequently forgone (e. g. Algazi et al. 2001, Kopčo and Shinn-Cunningham 2003, 2008, Keyrouz and Diepold 2008, Breebaart et al. 2010).

This section addresses specific aspects of the equalization, especially focusing on the benefit and corresponding effort for different procedures. The influence of intra-individual differences in microphone and HP positions on the HPTFs, in other words the violation of assumptions 5 and 6, is discussed along with mathematical basics of the equalization procedure. Further examined are implications of using nonindividual HPTFs for the equalization filter design and aspects of the HP production spread.

The results presented in this section are relevant also in hearing research with conventional HP playback (cf. Völk 2011a). If multiple subjects are included in a listening experiment, the inter-individual variability of the HPTFs must be considered. Further, for studies with more than one experimental run, the intra-individual reproducibility of the HPTFs due to HP repositioning has to be taken into account (cf. section 3.2.4).

5.2.1 Reproducibility of the Headphone Position

According to assumption 6, identical HP positions for playback and HPTF measurement are implied in the theoretical derivation of BS given in chapter 4. In the following, the consequences arising if assumption 6 is not fulfilled are discussed for the three HPTF measurement methods taken into account: probe microphone measurement with the probe tube tips in the auditory canals close to the eardrums, blocked auditory canal entrance miniature microphone measurement, and AH measurement.

For conventional AH measurements, Ryan and Furlong (1995) found less intra-individual variability due to HP repositioning in the TF of in-ear versus circum-aural HPs[31]. While comparable values are expected also for human head measurements, probe microphone and blocked auditory canal entrance miniature microphone measurements in combination with in-ear HPs are for mechanical reasons virtually impossible. For this reason, especially circum- and supra-aural HPs, which are frequently used in BS applications, are taken into account (Møller et al. 1995a, Spikofski and Fruhmann 2001, Mackensen et al. 1998, Pellegrini et al. 2007). However, the theoretical framework also applies to in-ear HPs.

Supra-aural HPs are coupled to the soft and flexible pinna, whereas circum-aural models are coupled to the stiff skull, typically without contacting the pinna. The coupling to the head is expected to be more reproducible, especially for AHs with inaccurately modeled pinnae. Consequently, a higher reproducibility, which is according to section 2.4 less variability, of the HP transfer characteristics due to repositioning is expected for circum- versus supra-aural HPs. Furthermore, a higher reproducibility may also be expected for AH measurements compared to human head measurements since AHs remain perfectly still. However, a tactile control of the HP fit is possible for human subjects positioning the HPs themselves, which might increase the reproducibility (Møller et al. 1996c).

Measurements with the HP model or even the specimen actually employed appear necessary based on the results of Shaw (1966), who addressed the position reproducibility

[31] Beyer DT 100 (circum-aural), Sony Twin Turbo (in-ear), six repositionings, Head Acoustics mannequin

of five HP models[32]. Shaw concluded: *“The substantial differences between the five earphones at any given frequency encourages one to believe that appreciable reductions in intrasubject range could be attained by appropriate design changes.”*

Therefore, reproducibility measurements carried out in the context of this work with one specimen of each of three HP models frequently employed in BS[33] are discussed based on a literature review in the following. Results are presented for blocked auditory canal entrance miniature microphone and for AH measurements, which represent according to chapter 4 the preferable approaches to BS if human head playback is targeted, while the literature review covers all three measurement methods taken into account. The measurements were carried out and analyzed for both ears, but no asymmetrical effects occurred. For that reason, only the left ear results are shown.

The discussion especially includes group delay data since to the author’s best knowledge, earlier reports on HPTF group delay or phase data are restricted to the statement of McAnally and Martin (2002) that in their blocked auditory canal entrance HPTF measurements[34] *”the variability of the group delays [...] was considerably less than the minimum discriminable interaural time difference”*. Even if this fact holds true for the eardrum TFs of a specific HP model, it represents no sufficient HP selection criterion since a group delay variation can alter not only ITDs but also the temporal energy distribution, possibly leading to frequency dependent differences of the interaural level patterns, especially for unsteady signals (Preis 1982). Furthermore, typical group delay distortions of professional HPs have been shown to be audible (Laws and Blauert 1973), especially when playing back binaural recordings, primarily influencing sound color and localization (Genuit 1986).

Probe Microphone Headphone Transfer Functions Shaw (1966) measured the reproducibility of the HPTF magnitude for three circum- and two supra-aural HPs[31] using probe microphones with the probe tube tips in the auditory canals of ten human subjects some millimeters in front of the eardrums. Three measurements with intermediate HP repositioning on each subject resulted in average intra-individual ranges smaller than 1 dB in the frequency range below 2 kHz and up to some 5 dB at higher frequencies. The supra-aural HPs show increasing intra-individual magnitude spectrum ranges up to 3 dB also for frequencies below 500 Hz. Shaw’s results for circum-aural HPs are supported qualitatively by the intra-individual standard-deviations of the probe microphone measurements presented by Wightman and Kistler (1989a)[35], as well as qualitatively and quantitatively by the results of Pralong and Carlile (1996)[36]. The results of Pralong and Carlile allow further comparison of human head and AH measurements, suggesting a somewhat reduced magnitude spectrum reproducibility for human listeners. However, this result is possibly influenced by variations in the probe microphone positions (cf. section 5.2.2).

[32] Mine Noisefoe Mk II, Sharpe-Philips, Sharpe HA 10, Telephonics TDH 39, Beyer DT 48

[33] a) Sennheiser HD 800, b) Stax λ pro NEW, c) Sennheiser HD 650

[34] Sennheiser HD 520 II, Head Acoustics HMS II.3 artificial head, three human subjects

[35] Sennheiser HD 340 headphones, ten repositionings, one human subject

[36] Sennheiser 250 Linear headphones, six repositionings, one human and one custom-made artificial head

Artificial Head Headphone Transfer Functions Kulkarni and Colburn (2000) measured twenty times the AH TFs of supra-aural HPs[37] with intermediate HP repositioning. The standard deviation of the corresponding magnitude spectra globally increases with frequency up to some 5 dB. Peak values of almost 10 dB occur in the frequency regions of the dips in the magnitude spectrum. The dips are subject to changes in depth and frequency for different HP positions, with the amount of the changes globally increasing with frequency. Kulkarni and Colburn stated that this *"variability* [...] *appears to be common to the general class of circumaural/supra-aural headphones"*. However, this conclusion is not validated for circum-aural HPs by a reference or data given by Kulkarni and Colburn (2000). Contrarily, Pralong and Carlile (1996) concluded from repositioning circum-aural HPs six times on an AH[38] *"that the placement of this type of headphones on the model head is highly reproducible. In contrast, the reproducibility of recordings obtained for repeated placements of headphones of the supra-aural type (Realistic Nova 17) was less satisfying."* Consequently, both studies indicate a similar reproducibility of the AH magnitude spectra of supra-aural HPs, while the conclusions differ for circum-aural HPs. The smoothed standard deviations reported by Zhong et al. (2010), representing 30 repositionings of circum-aural HPs on an AH[39], confirm the general structure of the data reported by Pralong and Carlile (1996), assuming the smoothing reduces peaks by about 2 dB. The structure and magnitude of the data are supported further by Ryan and Furlong (1995)[40] and McAnally and Martin (2002)[41]. In all earlier studies, large variability values occurred in the frequency ranges around prominent dips in the magnitude spectrum.

In the course of this work, for each of three circum-aural HP specimens of different models[42] 50 AH HPTFs according to equation 4.29 were measured with intermediate HP repositioning. Since the auditory-adapted analysis (AAA) spectrogram of the AH HPTF shown in figure 4.6 suggests primarily spectral effectiveness of HPTFs, the discussion is, according to section 2.4, based on the transfer characteristics shown by figure 5.3.

Each row of figure 5.3 represents the transfer characteristics of one HP specimen. The gray contours indicate the single measurements, the black contours the arithmetic mean values of magnitude spectra and group delays. By definition, the TFs visualized by figure 5.3 include the characteristics of the AH, the microphones, and the measurement system (cf. equation 4.29). However, it is possible to assess the AH HPTF reproducibility based on the data without loss of generality since the AH, the microphones, and the measurement system remained identical for all measurements.

The transfer characteristics show a comparable structure for all HP specimens, including the global increase of the magnitude spectrum variability with frequency and the variability maxima at the dips of the magnitude spectra reported earlier. Furthermore, a similar variability structure is visible also for the group delays. The group delay variability increases, apart from a maximum at the lower end of the HP transmission bandwidth,

[37] Sennheiser HD 520 headphones, KEMAR artificial head

[38] Sennheiser 250 Linear headphones, custom-made artificial head

[39] Sennheiser HD 250 II headphones, KEMAR artificial head

[40] Beyerdynamic DT 100, Head Acoustics mannequin, six repositionings

[41] Sennheiser HD 520 II headphones, Head Acoustics HMS II.3 artificial head, 20 repositionings

[42] a) Sennheiser HD 800, b) Stax λ pro NEW, c) Sennheiser HD 650, Neumann KU 80 artificial head

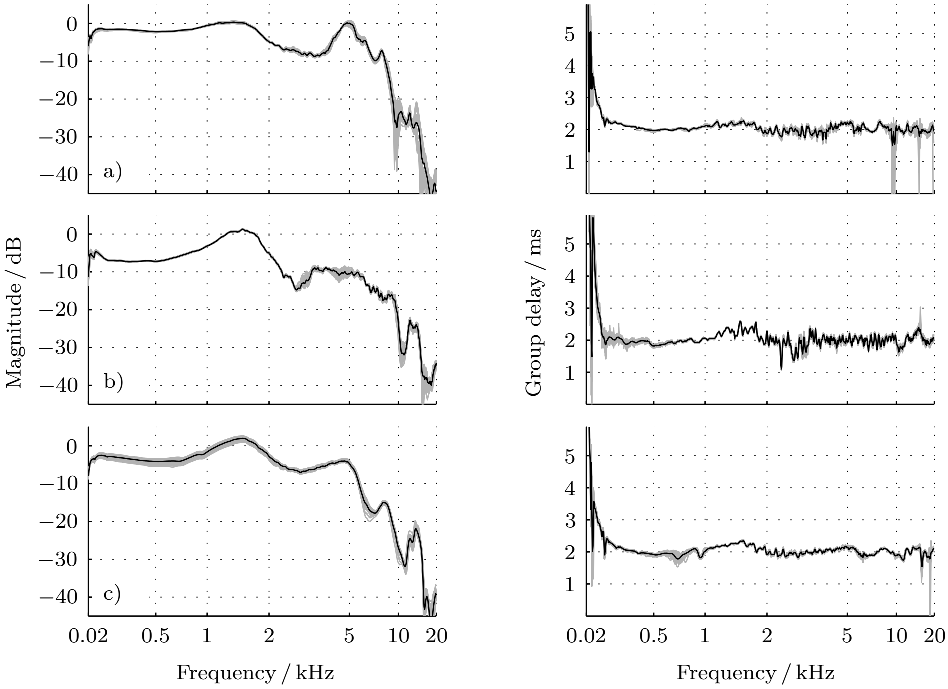

Figure 5.3: Artificial head transfer characteristics of three circum-aural headphones (a, b, c). Data from 50 measurements with headphone repositioning (gray contours) and arithmetic mean values (black contours).

globally with frequency, showing maxima at the dips of the magnitude spectra. In order to assess the reproducibility quantitatively, the variability percentiles $\mathrm{V}_{25}(S)$ and $\mathrm{V}_{75}(S)$ of the HP transfer characteristics according to section 2.4 are shown by figure 5.4.

Globally, structure and degree of the reproducibility are comparable for all three HP specimens. In the frequency range below a HP specimen dependent limiting frequency between 2 and 5 kHz, the magnitude spectrum variability proceeds frequency independently at small values. In the frequency range above the limiting frequency, a variability increase with frequency to about ±1 dB on average at the upper end of the audible frequency range is visible, exceeded by spurious maxima at frequencies corresponding to prominent dips of the magnitude spectra. In the frequency independent region, specimen c) shows some variability, presumably caused by its compressible ear pad, whereas the variability is almost negligible in the frequency independent range for specimens a) and b). However, specimen c) provides the widest frequency independent range. The group delay variability exhibits relatively high values at the lower limit of the HP transmission bandwidths, decaying for the different specimens more or less steeply towards the mid-frequency range. For higher frequencies, a global increase of the group delay variability is visible, with spurious peaks at frequencies roughly corresponding to prominent dips of the magnitude

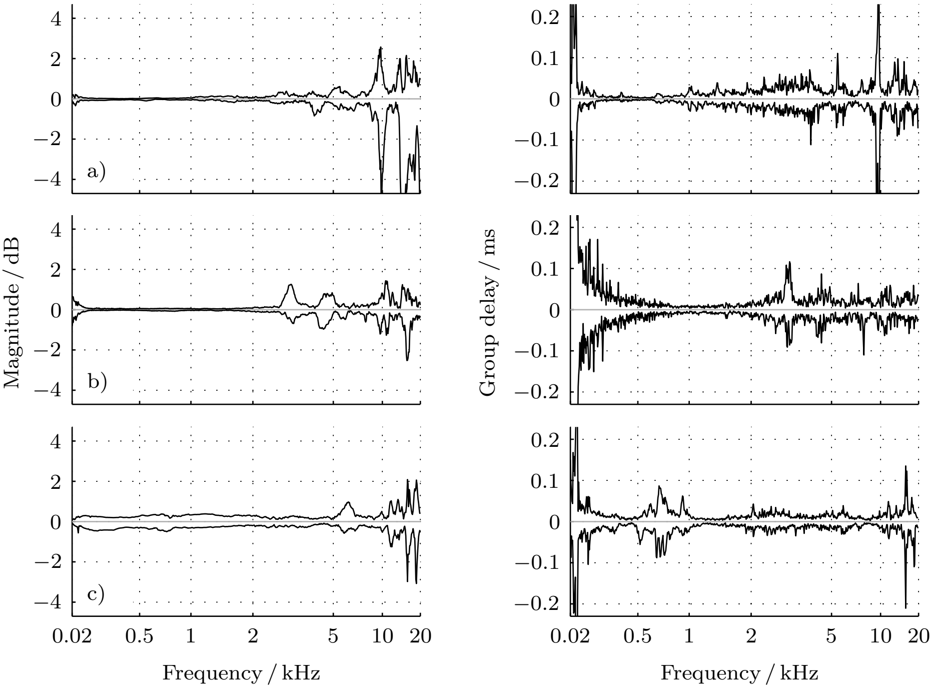

Figure 5.4: Artificial head reproducibility of the transfer characteristics of three circumaural headphones (a, b, c). 50 measurements with headphone repositioning.

spectra. On average, the group delay variability proceeds at some 20 μs. With regard to applications, the data shown provide an AH estimate of the reproducibility structure and amount to be expected for high quality studio HPs.

However, for being acquired with a randomly chosen specimen per model, the results are not representative for the corresponding model. For that reason, if detailed reproducibility values for a distinct HP specimen are of interest, measurements with the actual specimen using the procedures introduced are necessary.

Blocked Auditory Canal Headphone Transfer Functions Møller et al. (1995a) measured blocked auditory canal entrance miniature microphone HPTFs of human subjects, primarily concerned with the TFs as such, not the reproducibility. However, to check the validity of the results, Møller et al. repeated the measurements for each of twelve models[43] three times with the same subject. The repeatedly acquired magnitude spectra show a similar structure but differ in detail, comparable to the probe microphone and AH HPTFs discussed above. Also the standard deviations of the repeatedly measured spectrally smoothed magnitude

[43] Sony MRD-102, Jecklin float 2, Beyerdynamic DT 770, Beyerdynamic DT 990, Stax SR λ pro, Sennheiser 250 Linear, Sennheiser HD 420 SL, Sennheiser HD 540 reference, Sennheiser HD 560 ovation, AKG Acoustics K 240 DF, AKG Acoustics K 500, AKG Acoustics K 1000

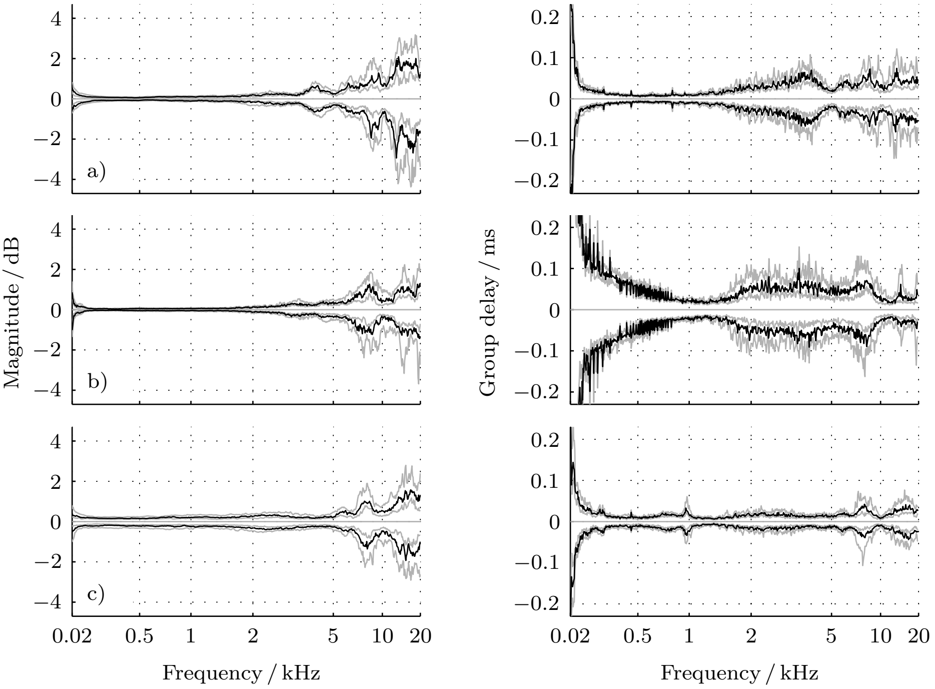

Figure 5.5: Reproducibility of the blocked auditory canal transfer characteristics of three circum-aural headphones (a, b, c). Inter-individual medians (black), 25% and 75% percentiles (gray, cf. section 2.4), data from 20 subjects, 50 measurements per subject.

spectra of circum-aural HPs reported by McAnally and Martin (2002)[44] as well as Zhong et al. (2010)[45] are structurally comparable to the probe microphone and AH situations. In detail, their standard deviations decay with frequency from about 1 dB at 100 Hz to almost 0 dB at 2 kHz, increasing again at higher frequencies, up to approximately 1 dB at 20 kHz, with spurious maxima in the range between 5 and 20 kHz.

For the three circum-aural HP specimens[46] also depicted by figures 5.3 and 5.4, 50 HPTFs[47] were measured according to equation 4.27 with intermediate HP repositioning at the entrances to the blocked auditory canals of each of 20 human subjects. The AAA spectrogram of a blocked auditory canal HPTF of specimen a), depicted by figure 4.3, suggests primary spectral effectiveness, which holds true also for specimens b) and c). Additionally considering section 2.4, the analysis is based on the HP transfer characteristics. Figure 5.5 shows the resulting reproducibility, each row representing one HP specimen. The black contours depict the inter-individual medians and the gray contours the 25% and

[44] Sennheiser HD 520 II, 20 repositionings on three human subjects
[45] Sennheiser HD 250 II, ten repositionings on one human subject
[46] a) Sennheiser HD 800, b) Stax λ pro NEW, c) Sennheiser HD 650
[47] Sennheiser KE 4-211-2 electret microphones in amplifier configuration mounted in foam earplugs

75% percentiles of the individual variability values $\mathrm{V}_{25}(S)$ and $\mathrm{V}_{75}(S)$ of the HP transfer characteristics according to section 2.4. The average human head reproducibility resembles the AH data shown by figure 5.4. Differences are the enlarged group delay variability of the human head blocked auditory canal entrance data, especially for specimen b), and by the inter-individual averaging globally smoothed characteristics. Overall, the variability characteristics tend to depend on the specimen. Thereby, the highest average magnitude spectrum variability occurs, due to the increased low-frequency variability, for specimen c), showing also the widest approximately frequency independent transmission bandwidth connected with the lowest group delay variability. Specimen b), the only model discussed working based on the electrostatic converter principle, shows the least magnitude spectrum variability but at the same time the highest group delay variation.

The increased group delay variability of the human head blocked auditory canal entrance measurements visible for HP specimen b) in figure 5.5 compared to the AH measurements shown by figure 5.4 may either be due to the transition from an artificial to human heads or due to the blocked auditory canal entrance measurement (cf. section 5.2.2). Addressing the blocked auditory canal entrance versus the standard AH measurement situation, figure 5.6 shows for HP specimen b) the blocked auditory canal AH transfer characteristics[48] according to equation 4.27 (black contours) and the corresponding conventional AH data redrawn from figure 5.4 b) for comparison purposes (gray contours).

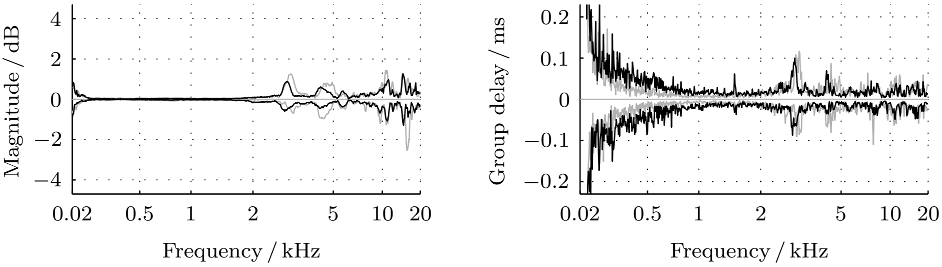

Figure 5.6: Artificial head reproducibility of the transfer characteristics of a circum-aural headphone (b) at the entrance to the blocked auditory canal (black contours) and at the eardrum (gray contours). Data based on 50 measurements with headphone repositioning.

The right panel of figure 5.6 indicates in the frequency range below about 3 kHz an enlarged group delay variability of the blocked auditory canal recording versus the standard AH recording, possibly influenced by the microphone position reproducibility discussed in the following section. It can be concluded from the comparison given by figure 5.6 that the increase of the group delay variability for specimen b) in figure 5.5 versus figure 5.4 is at least partly due to the transition from conventional AH to blocked auditory canal entrance measurements. The conclusion appears valid since the increase in group delay variability is shown by figure 5.6 to occur even in the AH situation.

[48] Sennheiser KE 4-211-2 electret microphones in amplifier configuration mounted in a rubber model ear

5.2.2 Reproducibility of the Microphone Position

According to assumption 5, the microphone positions for BS implementations are identical during recording and HPTF measurement, which is not necessarily valid. Wightman and Kistler (1989a) assessed the influence of repositioning a *probe microphone* tube tip in the auditory canal some millimeters in front of the eardrum on the TF from the free field to the probe microphone. The resulting standard deviation of the magnitude spectrum increases with frequency qualitatively comparable to the variability introduced by HP repositioning discussed in section 5.2.1. While Wightman and Kistler's results are not directly applicable to the TFs from HPs to probe microphones, they indicate a tendency for the expected variability structure.

Zhong et al. (2010) concluded, based on AH blocked auditory canal entrance measurements of the TFs of circum-aural HPs[49], the *miniature microphone* repositioning to be negligible if carried out by an experienced operator. However, in contrast to this conclusion, the smoothed standard deviations reported by Zhong et al. indicate considerable magnitude spectrum variation, especially in the frequency range of the most prominent dips of the magnitude spectrum.

For quantifying the variation of human head blocked auditory canal entrance HPTFs due to microphone repositioning, four sets of 50 HPTFs of specimen b)[50] were acquired in the context of this thesis for each of two human subjects on different days with intermediate microphone repositioning. In addition, the HPs were repositioned between every two measurements and the results were computed as the average of magnitude spectra and group delays of the 50 measurements. This procedure provides an average representation for each of the four sets, which can be regarded as the typical result for the respective microphone position, not influenced by the HP position variability. To acquire an overall representation of the four sets, $\mathrm{V}_{25}(S)$ and $\mathrm{V}_{75}(S)$ are computed according to section 2.4 individually. Figure 5.7 shows the resulting inter-individual average of the two subjects.

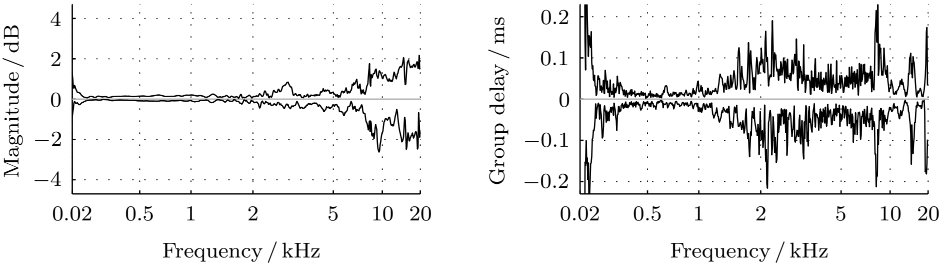

Figure 5.7: Human head reproducibility of blocked auditory canal headphone transfer characteristics due to microphone repositioning for a circum-aural headphone (b). Average result of two subjects based on four data sets with intermediate microphone repositioning per subject, each set averaged over 50 measurements with headphone repositioning.

[49] Sennheiser HD 250 II, KEMAR artificial head, comparison of two series of ten measurements

[50] Stax λ pro NEW

The variability introduced by different miniature microphone positions shown in figure 5.7 resembles qualitatively and quantitatively the variability due to the HP repositioning shown in figure 5.5 b). However, the variability due to microphone repositioning shown represents a worst case scenario with the microphones removed completely and repositioned afterwards. Therefore, the variability depicted exceeds the variability expected for microphone position variations during a measurement session. This conclusion is supported by a comparison to the differences between the black and gray contours in figure 5.6, revealing less variability due to microphone position variations for 50 blocked auditory canal entrance AH measurements than suggested by figure 5.7.

Since in contrast to human heads, AHs remain still during a measurement set, the variability for human head blocked auditory canal entrance measurements is expected in the range between the data shown by figure 5.7 and the difference of the gray and black contours in figure 5.6. Taking into account the data of figure 5.7 represent the worst case scenario of removing the microphones completely and repositioning them, these data can be considered an upper limit, unlikely reached in a carefully conducted measurement. This assumption is further supported by the considerably smaller average variability of human head blocked auditory canal HPTFs shown by figure 5.5, which include in addition to possibly occurring microphone position effects the variability due to the HP repositioning.

Therefore, the microphone position variability of carefully conducted individual BS implementations is expected closer to the AH situation than to the worst case situation shown in figure 5.7. The variability expected for the AH situation is described by the difference between the gray and black contours in figure 5.6, which is mainly a low-frequency group delay difference since the different frequencies of the variability maxima are due to different resonant frequencies in the open and blocked auditory canal situations and therefore not relevant for the discussion of the variability magnitude. That being said, figure 5.5 is assumed to show a good approximation of the reproducibility due to the HP repositioning, presumably influenced by microphone position effects only regarding low-frequency group delay values. However, it is not possible to fully rule out high frequency microphone position effects on magnitude spectrum and group delay. An upper limit is given for the magnitude spectrum by the smallest average values of figure 5.5 b) and regarding the group delay by the smallest average values of figure 5.5 c).

5.2.3 Inversion Problems in Binaural Synthesis

According to equation 4.43, the target for the BS equalization filter design is given by the inverted non-equalized BS situation TFs, which is possible since these TFs are invertible based on assumption 3. While assumption 3 can be valid in practical situations, it does not apply if the TF to be inverted shows zeros or small absolute values close to zero (notches in the magnitude spectrum or broader low-energy regions, as typically occurring at the limits of an electroacoustic transducer's transmission bandwidth, cf. Mourjopoulos et al. 1982, Preis 1982, Mourjopoulos 1988, Greenfield and Hawksford 1991, Potchinkov 1998a,b, Wightman and Kistler 2005).

In order to obtain a filter useful for practical implementations, amplification requirements of the filter target not provided by the audio processing equipment can be reduced by high- and low-pass filtering combined with frequency dependent regularization, which is the selective attenuation of undesired peaks (Kirkeby and Nelson 1998, 1999, Kirkeby et al. 1998, 1999, Norcross et al. 2006). Norcross et al. (2004b) showed by comparative listening experiments that only meticulously implemented inversion algorithms actually allow for the intended equalization while not introducing audible artifacts. The results of Norcross et al. further indicate that complex third-octave wide spectral smoothing prior to the inversion tends to reduce the audibility of artifacts, while for the test set not degrading the equalization result (cf. Norcross et al. 2002, 2003a,b). In addition, complex third-octave wide spectral smoothing temporally broadens the IR less than regularization (Norcross et al. 2004b). Recent tendencies aim at adapting the inversion to the audio signal to be processed (Norcross et al. 2005). In this case, computational complexity may become an important design criterion (Irisawa et al. 1998, Mouchtaris et al. 1999, 2000).

As a complex spectral smoothing approach adapted to the human hearing system, the AAS proposed in section 5.1.5 represents a perceptually motivated alternative to the physically motivated third-octave wide smoothing (Völk et al. 2011a). For that reason, AAS in combination with high- and low-pass filtering applied to the BS equalization target represents the preferable approach to BS equalization filter design.

5.2.4 Nonindividual Headphone Transfer Functions

The theoretically correct approach to BS is, according to chapter 4, individual synthesis, indicated by the superscript *ind*. In case individual recording or individual equalization is not desired or possible, completely or partially nonindividual procedures can be used (cf. definition 27). Regarding the equalization filter design, it is further possible to take into account only the magnitude spectrum of the individual equalization target, given by the inverted non-equalized BS situation TFs (cf. equation 4.43). This procedure is referred to as individual magnitude equalization (superscript *ind,abs*). With the real-valued parameterization factor c_p, linear phase individual magnitude equalization filters can be obtained based on equation 4.43 by

$$\mathbf{H}_\mathrm{eq}^\mathrm{ind,abs}(\mathbf{x}_{\mathrm{hp}_\mathrm{hptf}}, \mathbf{x}_{\mathrm{mic}_\mathrm{hptf}}) = \left|\mathbf{H}_\mathrm{eq}^\mathrm{ind}(\mathbf{x}_{\mathrm{hp}_\mathrm{hptf}}, \mathbf{x}_{\mathrm{mic}_\mathrm{hptf}})\right| \, \mathrm{e}^{-\,\mathrm{j}2\pi c_\mathrm{p} f/f_\mathrm{s}}. \tag{5.22}$$

Nonindividual magnitude equalization (superscript *nind,abs*) is defined in an analogous manner based on a nonindividual equalization filter target.

Human outer ear TFs show extreme values at frequencies typical for the specific subject (Mehrgardt and Mellert 1977). Consequently, averaging different individual equalization targets results in a loss of individual characteristics. However, if the average is computed with the aim of providing a BS equalization target which can improve the synthesis for arbitrary subjects, it is intended to include only the transfer characteristics common to most subjects. Regarding the procedure, it is important not to carry out the HPTF averaging in the time domain since different head and ear geometries may result in different first wave front arrival times, deteriorating time domain averaging, especially at high frequencies (Møller 1992). Therefore, typical filter design methods are based on averaging the individual magnitude spectra and combining the results either with a linear phase response or with the separately averaged phase data. Linear phase average magnitude equalization filter design (superscript *avg,abs*) based on averaging different individual magnitude spectra (superscript *avg*) is formulated independent of a specific procedure by

$$\mathbf{H}_{\mathrm{eq}}^{\mathrm{avg,abs}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{hptf}}}, \mathbf{x}_{\mathrm{mic}_{\mathrm{hptf}}}) = \left|\mathbf{H}_{\mathrm{eq}}^{\mathrm{ind}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{hptf}}}, \mathbf{x}_{\mathrm{mic}_{\mathrm{hptf}}})\right|^{\mathrm{avg}} \mathrm{e}^{-\,\mathrm{j}2\pi c_{\mathrm{p}} f/f_{\mathrm{s}}}. \tag{5.23}$$

For every equalization target, the resulting filters must be processed further by regularization or smoothing according to section 5.2.3 if the resulting dynamic range exceeds the dynamic range of the BS system. In the following, the consequences of nonindividual equalization procedures on the BS results are discussed for filter design based on probe microphone, AH, and blocked auditory canal entrance HPTFs separately. For the interpretation of probe and miniature microphone measurements, possibly occurring microphone position variability must be taken into account (cf. sections 5.1.4 and 5.2.2).

Probe Microphone Headphone Transfer Functions Pralong and Carlile (1996) studied inter-individual differences in the magnitude spectra of probe microphone HPTFs[51], finding the *"variability in the transfer functions was considerable for frequencies above 6 kHz [. . .]. In particular, the frequency and depth of the first spectral notch varied between 7.5 and 11 kHz and -15 and -40 dB, respectively. There were also considerable individual differences in the amplitude and the center frequency of the high-frequency gain features."* The corresponding inter-individual standard deviation reaches maxima of 10 dB between 2 and 3 kHz and exceeds 15 dB between 5 and 10 kHz. Pralong and Carlile concluded that individual equalization is required for BS implemented using probe microphone HPTFs, supported by measurements of Wightman and Kistler (2005)[52] and listening experiments of Kim and Choi (2005)[53], which reveal increased and thereby more correct horizontal plane externalization for individual versus no equalization.

Artificial Head Headphone Transfer Functions Kulkarni and Colburn (2000) studied the reproducibility of AH HPTFs[54] and assumed, based on their findings, an equalization target derived by averaging the results of multiple measurements, possibly on more than one subject, to be more useful in practical applications than an individual equalization target. According to Kulkarni and Colburn, an equalization filter based on a single measurement may not serve its purpose when the HPs are repositioned and introduce distortion. Using static BS systems with different HPs[55], implemented based on individual probe microphone recording at the eardrums, Yoshida et al. (2007) found in listening experiments an increased number of erroneous in-the-head localizations with equalization filters designed based on AH HPTFs compared to equalization filters designed based on individual probe microphone HPTFs. However, the amount of horizontal localization errors was reduced compared to the non-equalized situation for both equalization methods.

[51] Sennheiser 250 Linear headphones, ten subjects

[52] Beyerdynamic DT 990 headphones, five subjects

[53] Sennheiser HD 600 headphones, static binaural synthesis

[54] Sennheiser HD 520, KEMAR artificial head

[55] Sony MDR-ED 238, Sony MDR-E 737, Sennheiser HDA 200 headphones

In general, apart from differences in material properties, an AH may be regarded as a specific individual head. This holds especially true for blocked auditory canal entrance measurements since in this case the eardrum impedance does not influence the measurement results. Compared to nonindividual human head probe microphone measurements, it is further advantageous to be able to use measurement microphone capsules at the eardrum positions of AHs, enabling measurement results with increased signal to noise ratio and more realistic points of measurement in the auditory canals.

Blocked Auditory Canal Headphone Transfer Functions Møller et al. (1995a) measured the blocked auditory canal entrance HPTFs of 14 HP specimens[56] on 40 human subjects. The resulting magnitude spectra show inter-individual fluctuations of some 2 dB at low frequencies and up to 15 dB at high frequencies, primarily caused by resonances with individually different magnitude and frequency. Møller et al. conclude by inspecting the results that individual equalization is preferable because of the inter-individual differences, while average equalization may be acceptable, too. These conclusions are, however, not verified by Møller et al. (1995a). Aiming at verifying the aforementioned statement, providing group delay data in addition, and making data for additional HPs available, the inter-individual variability of the transfer characteristics of three circum-aural HP specimens[57] is addressed in the following, along with the equalization achievable with different equalization filter design approaches. For each of 20 subjects, 50 HPTFs per specimen were measured[58], according to equation 4.27, with intermediate HP repositioning. The individual transfer characteristics were computed by averaging magnitude spectra and group delays separately. This analysis based on the auditory-adapted transfer characteristics is justified according to section 2.4 by the AAA spectrograms of the HPTFs. Figure 4.3, showing the AAA spectrogram representing HP specimen a), suggests primarily spectral effectiveness of the HPTF, which holds true also for the AAA spectrograms of specimens b) and c). Figure 5.8 shows the individual blocked auditory canal entrance HP transfer characteristics (gray) and their arithmetic mean values (black), each row representing one HP specimen.

Regarding the magnitude spectra, the results of figure 5.8 confirm the data of Møller et al. (1995a) qualitatively by also showing a global variability increase over frequency and extreme values at individually different frequencies with varying magnitudes. All three specimens show group delay characteristics with a prominent global maximum at the lower limit of the HP transmission bandwidth, and with local extreme values corresponding to dips of the magnitude spectra. The inter-individual group delay variability also increases towards higher frequencies, with maxima roughly corresponding to dips of the magnitude spectrum. In order to address the variability not only qualitatively, but also quantitatively, figure 5.9 shows the inter-individual variability statistics corresponding to the transfer characteristics shown by figure 5.8, computed according to section 2.4.

[56] Sony MDR-102, Jecklin float model two, Beyerdynamic DT 770, Beyerdynamic DT 990, Stax SR λ pro, Sennheiser 250 Linear, Sennheiser HD 420 SL, Sennheiser HD 540 reference, Sennheiser HD 560 ovation, AKG Acoustics K 240 DF, AKG Acoustics K 500, AKG Acoustics K 1000

[57] a) Sennheiser HD 800, b) Stax λ pro NEW, c) Sennheiser HD 650

[58] Sennheiser KE 4-211-2 electret microphones in amplifier configuration mounted in foam earplugs

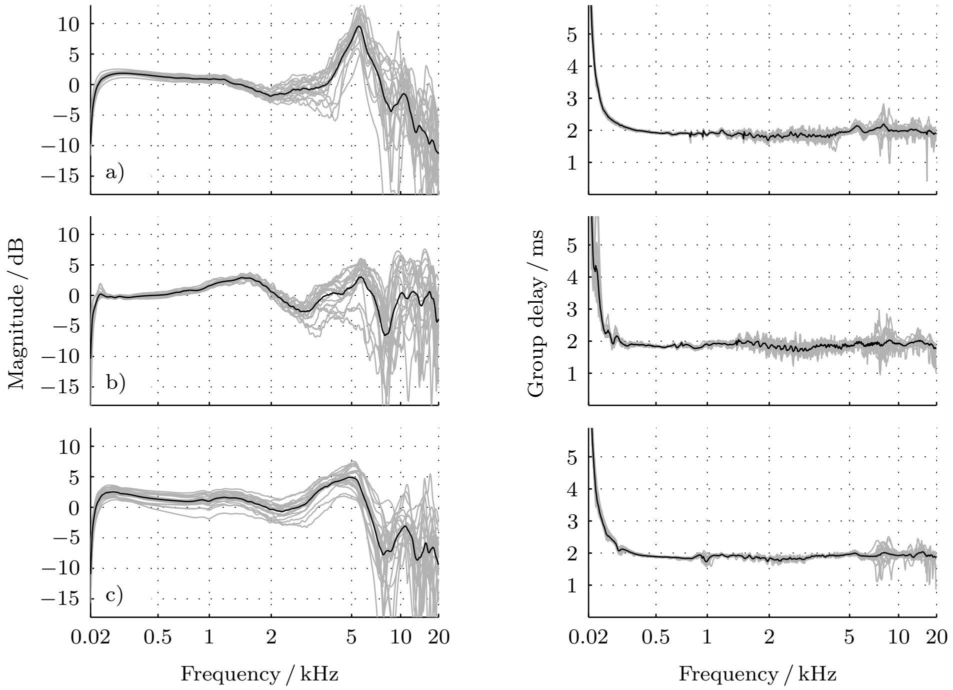

Figure 5.8: Human head blocked auditory canal transfer characteristics of three circum-aural headphones (a, b, c). Individual averages of 50 measurements with headphone repositioning for 20 subjects (gray contours) and inter-individual average (black contours).

The blocked auditory canal entrance HPTF magnitude spectrum variability characteristics depicted by the left column of figure 5.9 confirm the results reported by Møller et al. (1995a) quantitatively. All magnitude variability characteristics show a rather constant low-frequency region with variability values up to 1 dB, dependent on the respective HP specimen. At higher frequencies, the magnitude spectrum variability values increase up to more than 10 dB, with maxima at frequencies corresponding to extreme values of the magnitude spectra. The group delay variability proceeds structurally comparable to the magnitude spectrum variability, apart from a steep variability increase towards the lower limit of the audible frequency range and from a smaller nearly frequency independent region. The tendency for HP specimen c) to exhibit higher magnitude spectrum variability values in the frequency independent region than the other specimens, discovered in the HP position reproducibility shown by figure 5.5, is visible also in the inter-individual variability shown by figure 5.9. Overall, specimen c) provides the least group delay variability of the three specimens, at the expense of some ±0.5 dB enlarged magnitude spectrum variability.

The effort of implementing BS varies depending on the selected equalization method. In order to provide a basis for comparing effort and achievable results, the performance of the theoretically suboptimal methods individual magnitude equalization according to

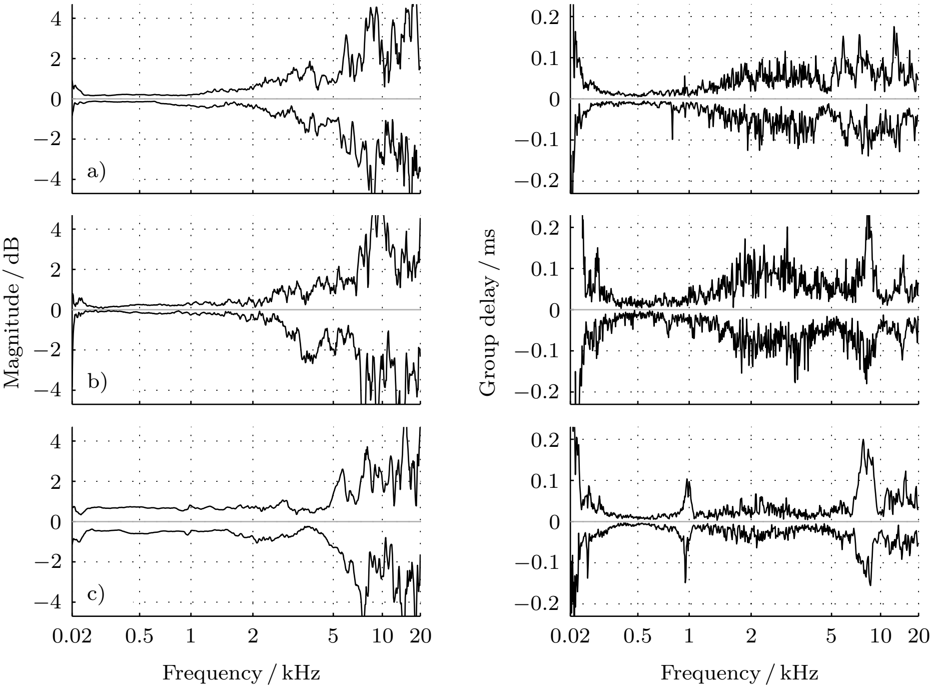

Figure 5.9: Inter-individual variability of the blocked auditory canal transfer characteristics of three circum-aural headphones (a, b, c). Data from 20 subjects individually averaged over 50 measurements with headphone repositioning.

equation 5.22 and average magnitude equalization according to equation 5.23 is addressed by the example of HP specimen b)[59]. Figure 5.10 shows individual blocked auditory canal entrance HP transfer characteristics, indicated by the gray contours, which are each computed by intra-individually averaging the magnitude spectra and group delays of 50 HPTFs measured with HP repositioning. The corresponding inter-individual averages are indicated by the black contours. In the first row the blocked auditory canal entrance HP transfer characteristics without equalization for 20 randomly selected human subjects (training set) are redrawn from figure 5.8 b). Based on these HPTFs, average magnitude equalization filters were designed according to equation 5.23 with directly using the blocked auditory canal entrance HPTFs according to equation 4.27 as the equalization targets. This condition is formulated based on equation 5.23 by setting

$$\mathbf{H}_{\mathrm{eq}}^{\mathrm{ind}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{hptf}}}, \mathbf{x}_{\mathrm{mic}_{\mathrm{hptf}}}) = \mathbf{H}_{\mathrm{hptf}_{\mathrm{m}}}^{\mathrm{ind,h,b}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{hptf}}}, \mathbf{x}_{\mathrm{m}_{\mathrm{hptf}}}). \tag{5.24}$$

These equalization targets are invalid targets for BS. However, the evaluation by HPTF measurement results for these targets in frequency independent magnitude spectra if

[59] Stax λ pro NEW, individual data averaged over 50 measurements with headphone repositioning

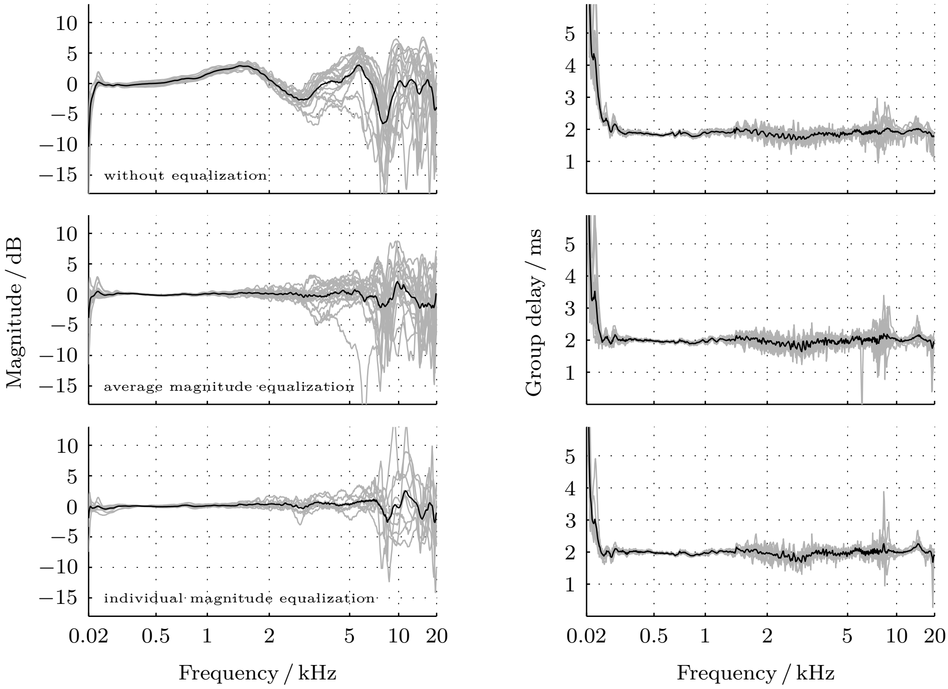

Figure 5.10: Blocked auditory canal human head transfer characteristics of a circum-aural headphone (b) with different equalization conditions (cf. inserts). Individual averages of magnitude and group delay over 50 measurements with headphone repositioning (gray contours) and inter-individual averages (black contours).

the equalization succeeds, simplifying the comparison of the approaches. Evaluating the average magnitude equalization according to equations 5.23 and 5.24 for ten subjects selected randomly from the training set and for ten others (test set) results in the data shown in the second row. Separate evaluations of the training and test sets revealed negligible differences to the overall results and are therefore not depicted. The third row represents the HP transfer characteristics measured with individual magnitude equalization according to equations 5.22 and 5.24 for ten randomly selected subjects.

Individual and average magnitude equalization increase the frequency independence of the blocked auditory canal entrance HPTF magnitude spectrum on average, while not substantially altering the group delay characteristics. Comparing the average magnitude equalization results in the middle row to the individual magnitude equalization results in the lower row reveals the effect of the individually different extreme values. Both equalization methods reduce, for most subjects, the depth of the resonances, while it is by definition impossible to achieve completely correct individual equalization with average magnitude equalization. However, also the individual magnitude equalization implemented here does not provide a perfect individual equalization. For some subjects, resonances are

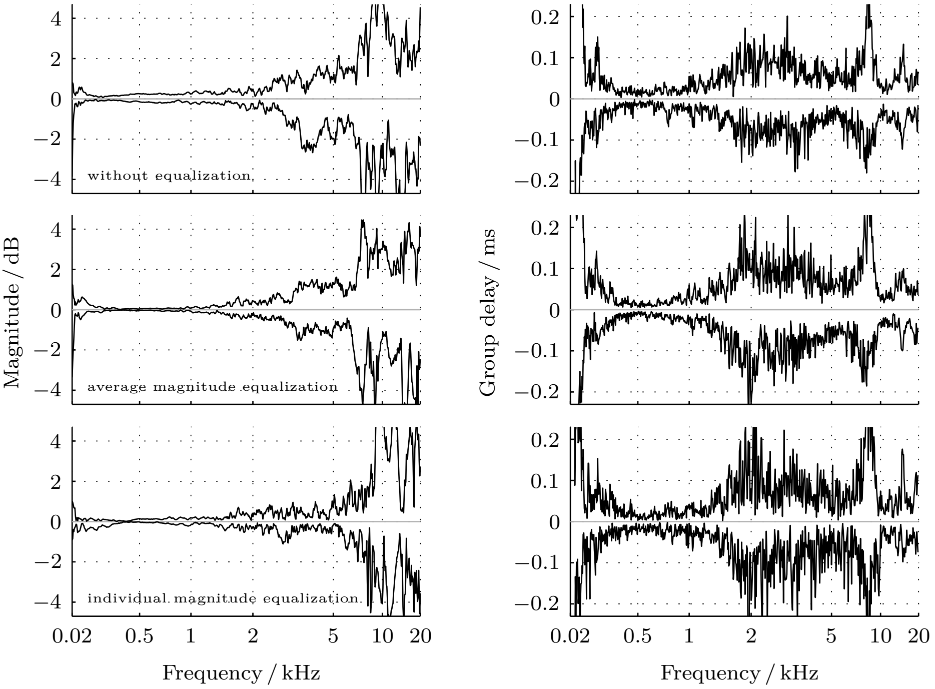

Figure 5.11: Variability of the blocked auditory canal human head transfer characteristics of a circum-aural headphone (b) with different equalization conditions (cf. inserts). Individual data averaged over 50 measurements with headphone repositioning.

overcompensated, resulting in narrow peaks and dips in the equalized magnitude spectra. The overcompensation can be reduced by iterative equalization procedures, which are not addressed in detail here (cf. Kim and Choi 2005).

In summary, the individual magnitude equalization according to equation 5.22 provides for all subjects a wide frequency independent magnitude spectrum range, while introducing narrowband artifacts for some subjects. The average magnitude equalization according to equation 5.23 enables a comparable average magnitude spectrum, while for most subjects a smaller frequency independent magnitude spectrum region and more pronounced extreme values occur. As intended, the group delay structure is preserved by both magnitude equalization approaches. In order to address the variability of the data depicted by figure 5.10 quantitatively, the corresponding inter-individual variability values are illustrated by figure 5.11.

Figure 5.11 proves that the inter-individual magnitude spectrum variability is reduced by the average and especially by the individual magnitude equalization, while the group delay variability is slightly increased globally, without changing its structure. The approximately constant frequency range of the magnitude spectrum variability is widened by the individual magnitude equalization, in contrast to the average magnitude equalization.

5.2.5 Headphone Production Spread

Wightman and Kistler (2005) found considerable differences between the magnitude spectra of probe microphone HPTFs of five specimens of the same circum-aural HP model[60] measured on the same listener. Based on these results, Wightman and Kistler postulated the need of designing BS equalization filters for the HP specimen actually used. In order to verify Wightman and Kistler's postulation, AH HPTFs of two specimens of each of two circum-aural HP models[61] were measured. Figure 5.12 shows the maxima of the *deviations within the models*, based on HPTFs averaged over 50 measurements with HP repositioning. The results were computed by taking, for each model, the differences of the magnitude spectra and group delays between the two HP specimens. The procedure was carried out separately for the left and right capsules, resulting in two sets of deviations (left and right) per model. Since two models were taken into account, the maxima shown by figure 5.12 were computed over four sets of deviations.

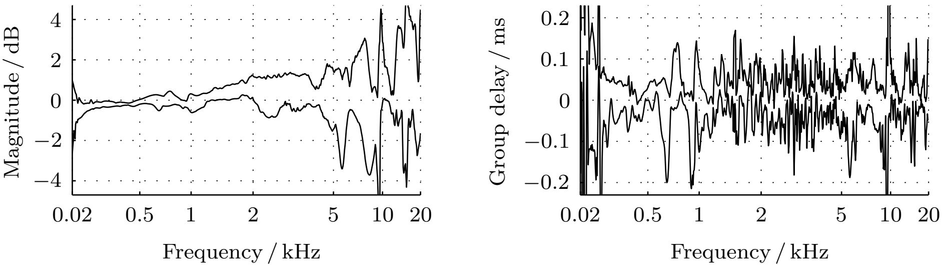

Figure 5.12: Maximum variability of the artificial head transfer characteristics of different specimens of the same circum-aural headphone model for two models. Data of each specimen averaged over 50 measurements with headphone repositioning.

In line with Wightman and Kistler (2005), the results shown by figure 5.12 indicate average intra-model differences of the HP transfer characteristics for the specimens studied in the range of the inter-individual variability. This finding may be interpreted in that measuring the BS equalization target with the HP specimen actually used can be considered as important as an individual equalization target. Based on the variability characteristics discussed in the present section, in addition to the inter-model differences of the transfer characteristics, inter-model differences of the reproducibility of the transfer characteristics were addressed. Regarding HP repositioning, the specimens of the same models provide qualitatively and quantitatively comparable reproducibility values (cf. section 5.2.1).

5.3 Headphone Selection for Blocked Auditory Canal Recording

Equations 4.53 and 4.71 in chapter 4 may be interpreted in that BS with blocked auditory canal entrance recording and blocked auditory canal entrance HPTF based equalization

[60] Beyerdynamic DT 990

[61] Sennheiser HD 800, Sennheiser HD 650; two models were included to increase the representativeness

possibly results in erroneous ear signals. It is shown in chapter 4 that the amount of error may be influenced by the specific HP model used. In section 5.4, perceptual consequences of the errors occurring when using inappropriate HPs are proven to be not compensable by level correction. For that reason, in this section the question is addressed whether HPs can be found that allow for the resulting ear signals to equal those of the reference scene. Appropriate HPs in this context preserve according to equations 4.53 and 4.71 the sound pressure transfer between the blocked auditory canal entrances and the eardrums. As a result, the blocked auditory canal headphone selection criterion (HPSC) is proposed, allowing the AH verification of the applicability of specific HPs for BS with blocked auditory canal entrance recording. The HPSC evaluates the applicability of HPs for BS with blocked auditory canal entrance recording, while other HP characteristics as for example the amount of nonlinear distortion or the human head transfer characteristics, important in conventional HP listening, are not addressed. For that reason, the HPSC provides no indication of the quality of HPs regarding conventional HP reproduction.

Based on a discussion of the existing approach of addressing the suitability of HPs for BS (Møller 1992), the HPSC is introduced and verified considering its implications on the overall BS TF. The HPSC is evaluated for two exemplary and randomly selected HP specimens, which are not representative of the respective HP models (cf. Völk 2012a).

5.3.1 Previous Headphone Selection Approach

Møller (1992) modeled the auditory canal by a transmission line, for a point source in the free field fully described by the open-circuit pressure spectrum P_2 and the radiation impedance Z_{ra} at the canal entrance. The pressure spectrum P_3 at the entrance is given using Thévenin's theorem (von Helmholtz 1853, Johnson 2003) to

$$P_3/P_2 = Z_{\mathrm{ec}}/\left(Z_{\mathrm{ec}} + Z_{\mathrm{ra}}\right), \tag{5.25}$$

splitting up the open-circuit pressure between the radiation impedance and the auditory canal impedance Z_{ec}. The transmission line model holds true for frequencies below some 10 kHz. In order to determine the open circuit-pressure, Møller used blocked auditory canal entrance miniature microphone measurement. For acquiring the pressure at the open entrance, he proposed probe microphone measurement, accepting the reduced signal to noise ratio compared to miniature microphones, which disturb the sound field in the auditory canal more. Assuming Z_{ra} to be independent of the source position, the TF from the point source spectrum P_1 to the eardrum spectrum P_4 is split in partial TFs by

$$P_4/P_1 = P_4/P_3 \cdot P_3/P_2 \cdot P_2/P_1. \tag{5.26}$$

For HP reproduction, Møller split up the transmission from the HP input voltage spectrum U_{hp} to the eardrum pressure spectrum P_7 using the pressure spectrum P_6 at the entrance to the auditory canal and its open-circuit counterpart P_5 by

$$P_7/U_{\mathrm{hp}} = P_7/P_6 \cdot P_6/P_5 \cdot P_5/U_{\mathrm{hp}}. \tag{5.27}$$

The transmission from P_5 to P_6 is given based on the transmission line model using Z_{hp}, the radiation impedance at the entrance to the auditory canal for HP playback, by

$$P_6/P_5 = Z_{\mathrm{ec}}/\left(Z_{\mathrm{ec}} + Z_{\mathrm{hp}}\right). \tag{5.28}$$

Z_{hp} is influenced by the characteristics of the volume enclosed by the HP and its mechanical and electrical subsystems. The transmission line model suggests identical transmission from the canal entrance to the eardrum for free-field and HP listening, formulated by

$$P_7/P_6 = P_4/P_3. \tag{5.29}$$

The relation of equations 5.25 and 5.28, describing the transfer between the sound pressure at the entrance to the auditory canal and its open-circuit counterpart, is given by

$$\left(P_3/P_2\right)/\left(P_6/P_5\right) = \left(Z_{\mathrm{ec}} + Z_{\mathrm{hp}}\right)/\left(Z_{\mathrm{ec}} + Z_{\mathrm{ra}}\right). \tag{5.30}$$

This ratio was later (Møller et al. 1995a) referred to as *pressure division ratio (PDR)*. The PDR is approximately one if $Z_{\mathrm{hp}} \approx Z_{\mathrm{ra}}$ holds true, that is if the HPs do not remarkably alter the radiation impedance, or if $Z_{\mathrm{ec}} \gg Z_{\mathrm{hp}}$ and $Z_{\mathrm{ec}} \gg Z_{\mathrm{ra}}$ are fulfilled, meaning no load of Z_{ra} exists. According to Møller (1992), the latter holds true for frequencies below about 1 kHz. If the PDR approximately equals one, equation 5.30 can be simplified to

$$P_3/P_2 \approx P_6/P_5. \tag{5.31}$$

HPs fulfilling equation 5.31 were referred to as *open headphones* by Møller (1992). Later (Møller et al. 1995a) the term *free-air equivalent coupling to the ear (FEC)* was introduced. Møller (1992) concluded, based on equations 5.29 and 5.30, that for correct BS with blocked auditory canal entrance recording using a miniature microphone described by the TF $M_1 = U_{\mathrm{M}}/P_{\mathrm{M}}$, a correction filter showing the TF

$$G_{\mathrm{C}} = \frac{P_4/P_3}{P_7/P_6}\frac{P_3/P_2}{P_6/P_5}\frac{1}{M_1 P_5/U_{\mathrm{hp}}} = \frac{Z_{\mathrm{ec}} + Z_{\mathrm{hp}}}{Z_{\mathrm{ec}} + Z_{\mathrm{ra}}}\frac{1}{M_1 P_5/U_{\mathrm{hp}}} \tag{5.32}$$

is required. He stated that the *"equalizing filter should include extra terms, if recording is made outside a blocked ear canal* [...]. *The extra terms are not required, when an open headphone is used* [...]*."* This quote shows that Møller assumed the PDR defined by equation 5.30 can be equalized for non FEC HPs. This assumption is not necessarily true because Z_{hp} includes the transfer characteristics of the volume enclosed by the HP and the mechanical and electrical HP subsystems. Consequently, resonances in the HP-head system may affect or prevent equalization (Groh 1974, Cox and D'Antonio 1997).

Measurement Procedure In order to determine PDRs with the motivation of addressing the FEC compliance of different HP models, Møller et al. (1995a) recorded the IRs representing the transfer paths from the input voltage spectrum U_{ls} of an LS to the pressure spectra $P_5^{'}$ and $P_6^{'}$ at probe microphones with the TFs M_2. These IRs are transformed to the corresponding TFs

$$H_5 = P_5^{'}/U_{\mathrm{ls}} = (P_5 M_2)\,/U_{\mathrm{ls}} \quad \text{and} \quad H_6 = P_6^{'}/U_{\mathrm{ls}} = (P_6 M_2)\,/U_{\mathrm{ls}}. \tag{5.33}$$

Møller et al. (1995a) reported that even though *"efforts were undertaken* [...]*, a small leak could arise between the headphone cushion and the head surface* [..., and the] *presence of the probe tube could also cause minor changes in the position and orientation of the headphone capsule"*. For these reasons, the same probe microphone was employed for both measurements, assuming *"this way the capsule displacement and the leak would have the same influence on P_5 and P_6, and the influence on the pressure division was eliminated."* In addition, the measurement sequence was selected carefully, first recording P_5 with the probe microphone in front of the blocked auditory canal entrance, then removing the earplug with *"as little disturbance of the probe microphone as possible"*, and recording P_6 afterwards. Based on the results and the free-field pressure divisions P_3/P_2 reported for the same subjects and the same equipment by Møller et al. (1995b), Møller et al. (1995a) computed according to equation 5.30 the PDRs

$$(P_3/P_2)\,(H_5/H_6) = (P_3/P_2)\,/\,(P_5/P_6)\,. \tag{5.34}$$

Møller et al. (1995a) stated that *"small changes in microphone and headphone positions between measurements with open and blocked ear canals and between free-air and headphone measurements* [...] *make*[*s*] *PDRs unreliable above approximately 7 kHz and thus they are not reported"*. However, Møller et al. addressed the question whether specific HPs show FEC characteristics based on the PDRs, neglecting spectral contributions above 7 kHz.

Shortcomings of the Approach Based on a transmission line model, the procedure proposed by Møller (1992) is in principle valid only if the wavelength is large compared to the auditory canal diameter. Møller assumed for typical parameters an upper limiting frequency of approximately 10 kHz. Above this frequency, equal sound pressure transfer from the auditory canal entrance to the eardrum in the HP and free-field situations is not necessarily given. For measuring the PDR, probe microphones are used, showing reduced signal to noise ratio compared to miniature microphones (Møller et al. 1995a). Further, Møller's procedure requires the assumption of Z_{ra} to be independent of the source position, which is not necessarily fulfilled (Hammershøi and Møller 1996a,b). However, the most significant shortcoming is of a procedural nature: Assuming the influences of the probe microphones on the HPTFs during the measurements of P_5 and P_6 with intermediate HP repositioning and earplug removal remain constant and therefore cancel each other when computing the PDR is questionable. Comparable reproducibility issues evolve regarding the probe microphone tube tip positions during the measurements of P_5 and P_6. These procedural shortcomings are especially likely to result in high frequency errors, while probably also causing leakage effects at the lower limit of the HP transmission bandwidth. Questioning the validity of the results especially at high frequencies is supported by the decision of Møller et al. (1995b) not to report the PDRs at frequencies above 7 kHz.

The limitation of correct BS to the audible frequency range below 7 kHz is auditory relevant, especially considering the relevance of high frequency information for elevation localization (Asano et al. 1990, Middlebrooks and Green 1991, Wightman and Kistler 1997) and sharpness perception (von Bismarck 1971, Fastl and Zwicker 2007, pp. 239–241). Additionally, the sharpness is considered a major influence factor on auditory pleasantness (Aures 1984) and timbre (von Bismarck 1974). Therefore, a selection criterion covering the full audible frequency range is introduced and verified in the remainder of this section.

5.3.2 Headphone Selection Criterion

Based on the BS theory introduced in chapter 4, the requirements for HPs to be appropriate for BS with blocked auditory canal entrance recording and equalization based on blocked auditory canal entrance HPTFs without further equalization can be derived from equation 4.53. Mathematically, for appropriate HPs, the TFs

$$\mathbf{H}^{\mathrm{ind}}_{\mathrm{hpsc}}(\mathbf{x}_{\mathrm{m}_{\mathrm{rec}}},\mathbf{x}_{\mathrm{hp}_{\mathrm{play}}},\mathbf{x}_{\mathrm{hp}_{\mathrm{hptf}}},\mathbf{x}_{\mathrm{m}_{\mathrm{hptf}}}) = \frac{\mathbf{H}^{\mathrm{ind}}_{\mathbf{p}_{\mathrm{m}_{\mathrm{b}}},\mathbf{p}_{\mathrm{e}},\mathrm{ls}}(\mathbf{x}_{\mathrm{m}_{\mathrm{rec}}})}{\mathbf{H}^{\mathrm{ind,h}}_{\mathbf{p}_{\mathrm{m}_{\mathrm{b}}},\mathbf{p}_{\mathrm{e}},\mathrm{hp}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{play}}},\mathbf{x}_{\mathrm{hp}_{\mathrm{hptf}}},\mathbf{x}_{\mathrm{m}_{\mathrm{hptf}}})} \tag{5.35}$$

must frequency independently equal one. This requirement is defined as the blocked auditory canal headphone selection criterion (HPSC) here (subscript *hpsc*, cf. Völk 2010a, 2012a), formulated using equations 4.52 and 4.25 by

$$\begin{aligned}&\mathbf{H}^{\mathrm{ind}}_{\mathrm{hpsc}}(\mathbf{x}_{\mathrm{m}_{\mathrm{rec}}},\mathbf{x}_{\mathrm{hp}_{\mathrm{play}}},\mathbf{x}_{\mathrm{hp}_{\mathrm{hptf}}},\mathbf{x}_{\mathrm{m}_{\mathrm{hptf}}}) =\\ &\quad = \frac{\mathbf{H}^{\mathrm{ind}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{e}}}(\mathbf{x}_{\mathrm{h}_{\mathrm{ref}}},\mathbf{x}_{\mathrm{ls}_{\mathrm{ref}}})}{\mathbf{H}^{\mathrm{ind,b}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{m}}}(\mathbf{x}_{\mathrm{h}_{\mathrm{ref}}},\mathbf{x}_{\mathrm{ls}_{\mathrm{ref}}},\mathbf{x}_{\mathrm{m}_{\mathrm{rec}}})}\cdot\frac{\mathbf{H}^{\mathrm{ind,h,b}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{m}}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{hptf}}},\mathbf{x}_{\mathrm{m}_{\mathrm{hptf}}})}{\mathbf{H}^{\mathrm{ind,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{e}}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{play}}})} \overset{!}{=} \mathbf{1}.\end{aligned} \tag{5.36}$$

Equation 5.36 may be interpreted as follows: The relations of the sound pressure spectra at the eardrums to the sound pressure spectra at the entrances of the blocked auditory canals, according to chapter 4 referred to as the blocking factors, have to be identical for LS and HP reproduction. The theoretical optimum of equality is not reached by most measurement setups and HPs. Hence, application specific acceptability limits must be imposed. Further, the HPSC is valid only at frequencies where the LS provides sufficient energy for the measurements not to be disturbed by noise.

The LS and HP situations differ in the active sound source and in the fact that HPs are present in the HP playback and HPTF measurement situations, in contrast to the LS based recording situation and reference scene. Differences in the blocking factors for LS and HP reproduction may be due either to the different sources or to the mere HP presence. The contribution of the sources may be isolated by considering BS with the HP reference scene given by definition 30 (for the associated system theory cf. appendix C).

Definition 30 (*Headphone Reference Scene for Binaural Synthesis*)

The headphone reference scene for binaural synthesis consists of a subject wearing inactive headphones listening to a loudspeaker in a reverberant listening environment.

Using BS with HP reference scene, the blocking factors in the HP and LS situations are identical apart from the active sound sources. Therefore, the HPSC with HP reference scene predicts the sound source influence (indicated by the additional subscript s), and can be formulated assuming identical HP positions $\mathbf{x}_{\mathrm{hp_{ref}}} = \mathbf{x}_{\mathrm{hp_{rec}}}$ in the HP reference scene and the recording situation by

$$\begin{aligned}\mathbf{H}^{\mathrm{ind,h}}_{\mathrm{hpsc,s}}(\mathbf{x}_{\mathrm{m_{rec}}},\mathbf{x}_{\mathrm{hp_{play}}},\mathbf{x}_{\mathrm{hp_{hptf}}},\mathbf{x}_{\mathrm{m_{hptf}}}) &= \frac{\mathbf{H}^{\mathrm{ind,h}}_{\mathbf{p}_{\mathrm{m_b}},\mathbf{p}_{\mathrm{e}},\mathrm{ls}}(\mathbf{x}_{\mathrm{m_{rec}}},\mathbf{x}_{\mathrm{hp_{rec}}},\mathbf{x}_{\mathrm{hp_{ref}}})}{\mathbf{H}^{\mathrm{ind,h}}_{\mathbf{p}_{\mathrm{m_b}},\mathbf{p}_{\mathrm{e}},\mathrm{hp}}(\mathbf{x}_{\mathrm{hp_{play}}},\mathbf{x}_{\mathrm{hp_{hptf}}},\mathbf{x}_{\mathrm{m_{hptf}}})} \\ &= \frac{\mathbf{H}^{\mathrm{ind,h}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{e}}}(\mathbf{x}_{\mathrm{h_{ref}}},\mathbf{x}_{\mathrm{ls_{ref}}},\mathbf{x}_{\mathrm{hp_{ref}}})}{\mathbf{H}^{\mathrm{ind,b,h}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{m}}}(\mathbf{x}_{\mathrm{h_{ref}}},\mathbf{x}_{\mathrm{ls_{ref}}},\mathbf{x}_{\mathrm{m_{rec}}},\mathbf{x}_{\mathrm{hp_{rec}}})} \cdot \frac{\mathbf{H}^{\mathrm{ind,h,b}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{m}}}(\mathbf{x}_{\mathrm{hp_{hptf}}},\mathbf{x}_{\mathrm{m_{hptf}}})}{\mathbf{H}^{\mathrm{ind,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{e}}}(\mathbf{x}_{\mathrm{hp_{play}}})}.\end{aligned} \tag{5.37}$$

Combining the data acquired by the blocked auditory canal HP selection criteria with and without HPs, it is possible to isolate the contribution of the mere HP presence. Mathematically, the influence of the HP presence (additional subscript p) is given using equations 5.35 and 5.37 by

$$\mathbf{H}^{\mathrm{ind,h}}_{\mathrm{hpsc,p}}(\mathbf{x}_{\mathrm{m_{rec}}},\mathbf{x}_{\mathrm{hp_{play}}},\mathbf{x}_{\mathrm{hp_{hptf}}},\mathbf{x}_{\mathrm{m_{hptf}}}) = \frac{\mathbf{H}^{\mathrm{ind}}_{\mathrm{hpsc}}(\mathbf{x}_{\mathrm{m_{rec}}},\mathbf{x}_{\mathrm{hp_{play}}},\mathbf{x}_{\mathrm{hp_{hptf}}},\mathbf{x}_{\mathrm{m_{hptf}}})}{\mathbf{H}^{\mathrm{ind,h}}_{\mathrm{hpsc,s}}(\mathbf{x}_{\mathrm{m_{rec}}},\mathbf{x}_{\mathrm{hp_{play}}},\mathbf{x}_{\mathrm{hp_{hptf}}},\mathbf{x}_{\mathrm{m_{hptf}}})}. \tag{5.38}$$

5.3.3 Artificial Head Approximation

Equation 5.36 reveals that the evaluation of the HPSC with human heads involves ear signal measurements, which are in general not perfectly correct because an ear signal measurement in a strict sense would require acquiring the sound pressure detected by the eardrum (cf. section 5.1.4). However, the HPSC may be evaluated without loss of generality using an AH since the blocking factors are addressed, representing HP properties which are independent of the specific evaluation head (cf. section 5.3.4). If the AH is designed to allow for positioning the microphones *reproducibly* at the blocked auditory canal entrances and at the eardrum positions, equation 5.35 can be simplified to

$$\begin{aligned}\mathbf{H}^{\mathrm{ah}}_{\mathrm{hpsc}}(\mathbf{x}_{\mathrm{m_{rec}}},\mathbf{x}_{\mathrm{hp_{play}}},\mathbf{x}_{\mathrm{hp_{hptf}}},\mathbf{x}_{\mathrm{m_{hptf}}}) &= \frac{\mathbf{H}^{\mathrm{ah}}_{\mathbf{p}_{\mathrm{m_b}},\mathbf{p}_{\mathrm{ahm_e}},\mathrm{ls}}(\mathbf{x}_{\mathrm{m_{rec}}})}{\mathbf{H}^{\mathrm{ah,h}}_{\mathbf{p}_{\mathrm{m_b}},\mathbf{p}_{\mathrm{ahm_e}},\mathrm{hp}}(\mathbf{x}_{\mathrm{hp_{play}}},\mathbf{x}_{\mathrm{hp_{hptf}}},\mathbf{x}_{\mathrm{m_{hptf}}})} \\ &= \frac{\mathbf{H}^{\mathrm{ah}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{h_{ref}}},\mathbf{x}_{\mathrm{ls_{ref}}})}{\mathbf{H}^{\mathrm{ah,b}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{m}}}(\mathbf{x}_{\mathrm{h_{ref}}},\mathbf{x}_{\mathrm{ls_{ref}}},\mathbf{x}_{\mathrm{m_{rec}}})} \cdot \frac{\mathbf{H}^{\mathrm{ah,h,b}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{m}}}(\mathbf{x}_{\mathrm{hp_{hptf}}},\mathbf{x}_{\mathrm{m_{hptf}}})}{\mathbf{H}^{\mathrm{ah,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{hp_{play}}})} \overset{!}{=} \mathbf{1}.\end{aligned} \tag{5.39}$$

Based on equations 4.15, 4.17, 4.27, and 4.29, it is possible to rewrite equation 5.39 in a practically more applicable way, solely dependent on HP and recording situation TFs, by

$$\begin{aligned}&\mathbf{H}^{\mathrm{ah}}_{\mathrm{hpsc}}(\mathbf{x}_{\mathrm{m_{rec}}},\mathbf{x}_{\mathrm{hp_{play}}},\mathbf{x}_{\mathrm{hp_{hptf}}},\mathbf{x}_{\mathrm{m_{hptf}}}) = \\ &\qquad = \frac{\mathbf{H}^{\mathrm{ah}}_{\mathrm{rec_{ahm}}}(\mathbf{x}_{\mathrm{h_{ref}}},\mathbf{x}_{\mathrm{ls_{ref}}})}{\mathbf{H}^{\mathrm{ah,b}}_{\mathrm{rec_m}}(\mathbf{x}_{\mathrm{h_{ref}}},\mathbf{x}_{\mathrm{ls_{ref}}},\mathbf{x}_{\mathrm{m_{rec}}})} \cdot \frac{\mathbf{H}^{\mathrm{ah,h,b}}_{\mathrm{hptf_m}}(\mathbf{x}_{\mathrm{hp_{hptf}}},\mathbf{x}_{\mathrm{m_{hptf}}})}{\mathbf{H}^{\mathrm{ah,h}}_{\mathrm{hptf_{ahm}}}(\mathbf{x}_{\mathrm{hp_{play}}})}.\end{aligned} \tag{5.40}$$

In this form, for evaluating the suitability of HPs for BS with blocked auditory canal recording, the HPSC requires four measurements on an AH that enables positioning a microphone reproducibly at the eardrum position and at the blocked auditory canal entrance. Exemplary HPSC transfer characteristics acquired according to equation 5.40 using an AH as described above[62] are depicted in figure 5.13, where HP specimen a) is indicated by the black contours, specimen b) by the gray contours.

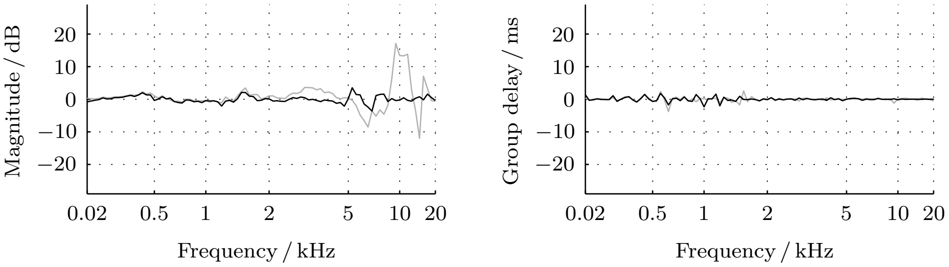

Figure 5.13: Blocked auditory canal headphone selection criterion evaluated for two different headphone specimens on an artificial head. Specimen a) is indicated by the black contours, specimen b) by the gray contours.

Figure 5.13 reveals frequency dependence of the HPSC magnitude spectrum and group delay for both HP specimens. While the magnitude spectrum of specimen a), indicated by the black contour, proceeds independent of frequency within ±3 dB, the HPSC magnitude spectrum of specimen b) exhibits spectral peaks exceeding 15 dB in the frequency range above about 5 kHz. Since the HPSC magnitude spectra are comparable in the frequency range below some 3 kHz, the frequency dependencies in this range are likely caused by the measurement setup itself. The HPSC group delays are, apart from the procedural effects, approximately frequency independent. Overall, specimen a) is indicated more appropriate for BS with blocked auditory canal entrance recording by the HPSC.

Analogue to equation 5.39, the HPSC with HP reference scene given by equation 5.38 can be adapted to the AH case. Assuming identical HP positions $\mathbf{x}_{\mathrm{hp}_{\mathrm{ref}}} = \mathbf{x}_{\mathrm{hp}_{\mathrm{rec}}}$ in the HP reference scene and the recording situation, the HPSC is therefore given by

$$\begin{aligned}\mathbf{H}^{\mathrm{ah,h}}_{\mathrm{hpsc,s}}(\mathbf{x}_{\mathrm{m}_{\mathrm{rec}}}, \mathbf{x}_{\mathrm{hp}_{\mathrm{play}}}, \mathbf{x}_{\mathrm{hp}_{\mathrm{hptf}}}, \mathbf{x}_{\mathrm{m}_{\mathrm{hptf}}}) &= \frac{\mathbf{H}^{\mathrm{ah,h}}_{\mathbf{p}_{\mathrm{m_b}},\mathbf{p}_{\mathrm{ahm_e}},\mathrm{ls}}(\mathbf{x}_{\mathrm{m}_{\mathrm{rec}}}, \mathbf{x}_{\mathrm{hp}_{\mathrm{rec}}}, \mathbf{x}_{\mathrm{hp}_{\mathrm{ref}}})}{\mathbf{H}^{\mathrm{ah,h}}_{\mathbf{p}_{\mathrm{m_b}},\mathbf{p}_{\mathrm{ahm_e}},\mathrm{hp}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{play}}}, \mathbf{x}_{\mathrm{hp}_{\mathrm{hptf}}}, \mathbf{x}_{\mathrm{m}_{\mathrm{hptf}}})} \\ &= \frac{\mathbf{H}^{\mathrm{ah,h}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{h}_{\mathrm{ref}}}, \mathbf{x}_{\mathrm{ls}_{\mathrm{ref}}}, \mathbf{x}_{\mathrm{hp}_{\mathrm{ref}}})}{\mathbf{H}^{\mathrm{ah,b,h}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{m}}}(\mathbf{x}_{\mathrm{h}_{\mathrm{ref}}}, \mathbf{x}_{\mathrm{ls}_{\mathrm{ref}}}, \mathbf{x}_{\mathrm{m}_{\mathrm{rec}}}, \mathbf{x}_{\mathrm{hp}_{\mathrm{rec}}})} \cdot \frac{\mathbf{H}^{\mathrm{ah,h,b}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{m}}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{hptf}}}, \mathbf{x}_{\mathrm{m}_{\mathrm{hptf}}})}{\mathbf{H}^{\mathrm{ah,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{play}}})} \\ &= \frac{\mathbf{H}^{\mathrm{ah,h}}_{\mathrm{rec}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{h}_{\mathrm{ref}}}, \mathbf{x}_{\mathrm{ls}_{\mathrm{ref}}}, \mathbf{x}_{\mathrm{hp}_{\mathrm{ref}}})}{\mathbf{H}^{\mathrm{ah,b,h}}_{\mathrm{rec}_{\mathrm{m}}}(\mathbf{x}_{\mathrm{h}_{\mathrm{ref}}}, \mathbf{x}_{\mathrm{ls}_{\mathrm{ref}}}, \mathbf{x}_{\mathrm{m}_{\mathrm{rec}}}, \mathbf{x}_{\mathrm{hp}_{\mathrm{rec}}})} \cdot \frac{\mathbf{H}^{\mathrm{ah,h,b}}_{\mathrm{hptf}_{\mathrm{m}}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{hptf}}}, \mathbf{x}_{\mathrm{m}_{\mathrm{hptf}}})}{\mathbf{H}^{\mathrm{ah,h}}_{\mathrm{hptf}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{play}}})} .\end{aligned} \tag{5.41}$$

[62] a) Sennheiser HD 800, b) Stax λ pro NEW headphones, Klein + Hummel Studio Monitor Loudspeaker O 98, custom-made artificial head AH$_\mathrm{c}$

The influences of the HP presence on the HPSC are given for the AH scenario analogously to the corresponding human head situation (equation 5.38) by

$$\mathbf{H}^{\mathrm{ah,h}}_{\mathrm{hpsc,p}}(\mathbf{x}_{\mathrm{m_{rec}}}, \mathbf{x}_{\mathrm{hp_{play}}}, \mathbf{x}_{\mathrm{hp_{hptf}}}, \mathbf{x}_{\mathrm{m_{hptf}}}) = \frac{\mathbf{H}^{\mathrm{ah}}_{\mathrm{hpsc}}(\mathbf{x}_{\mathrm{m_{rec}}}, \mathbf{x}_{\mathrm{hp_{play}}}, \mathbf{x}_{\mathrm{hp_{hptf}}}, \mathbf{x}_{\mathrm{m_{hptf}}})}{\mathbf{H}^{\mathrm{ah,h}}_{\mathrm{hpsc,s}}(\mathbf{x}_{\mathrm{m_{rec}}}, \mathbf{x}_{\mathrm{hp_{play}}}, \mathbf{x}_{\mathrm{hp_{hptf}}}, \mathbf{x}_{\mathrm{m_{hptf}}})}. \tag{5.42}$$

Figure 5.14 shows the transfer characteristics describing the AH approximation of the HPSC with the HP reference scene according to equation 5.41, that is in other words purely the source effect. HP specimen a) is indicated by the black contours, HP specimen b) by the gray contours[63].

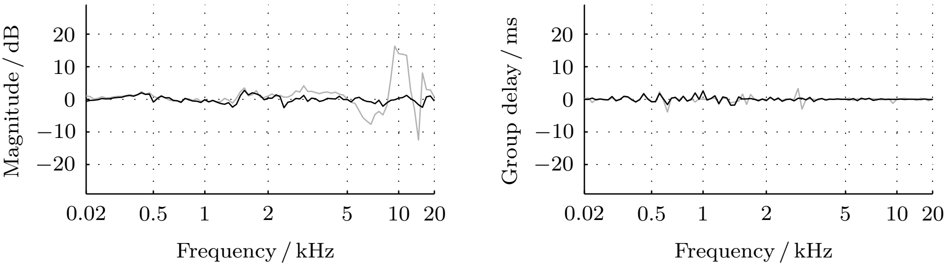

Figure 5.14: Blocked auditory canal headphone selection criterion with headphone reference evaluated for two different headphone specimens on an artificial head. Specimen a) is indicated by the black contours, specimen b) by the gray contours.

The results shown by figure 5.14 are qualitatively and quantitatively comparable to those given by figure 5.13. Consequently, for the HP specimens under consideration, the source effect represents the major influence factor on the artifacts occurring especially for specimen b). Therefore, specimen b) is assumed less suited for BS with blocked auditory canal entrance recording than specimen a). The validity of this prognosis is verified by means of loudness comparisons in section 5.4 (cf. Völk et al. 2011d, Völk and Fastl 2011a).

5.3.4 Stability, Repeatability, and Hardware Influences

In order to address the repeatability of the HPSC, seven measurements according to equation 5.40 were carried out with the same hardware configuration[64], but using different LS positions. Therefore, a possible influence of the sound incidence direction would be included in the results in addition to the variability representing the HPSC repeatability. Figure 5.15 shows the transfer characteristics describing each HPSC measurement, indicated by the gray contours, and the arithmetic mean values of magnitude spectra and group delays, represented by the black contours.

[63] a) Sennheiser HD 800, b) Stax λ pro NEW headphones, Klein + Hummel Studio Monitor Loudspeaker O 98, custom-made artificial head AH$_\mathrm{c}$

[64] Stax λ pro NEW headphones, Klein + Hummel Studio Monitor Loudspeaker O 200, custom-made artificial head AH$_\mathrm{c}$

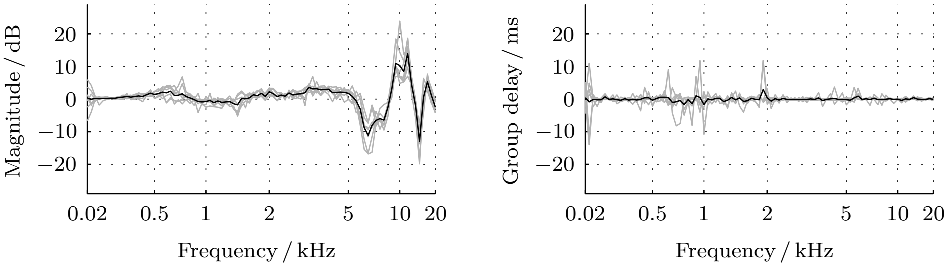

Figure 5.15: Blocked auditory canal headphone selection criterion evaluated for headphone specimen b) on an artificial head. The results are shown for different sound incidence directions (gray contours) and on average (black contours).

The results depicted by figure 5.15 reveal the existence of the prominent spectral characteristics in all measurement results, but at slightly different resonant frequencies. Therefore, the amount of the extreme values is somewhat reduced on average, compared to the single measurements. However, an approximate prognosis of the average characteristics from a single measurement is assumed valid based on HPSC measurements according to equation 5.40 for magnitude spectra and group delays. It can be concluded that possibly occurring influences of the sound incidence direction and the reproducibility of the HPSC lie within the accuracy of the prototypical measurement system employed.

Figure 5.16 shows the transfer characteristics of the AH HPSC according to equation 5.40, measured in a different room and with a different LS[65] compared to figure 5.13. In order to isolate the influence of the specific measurement situation, the same HPs were used. The similarity of figures 5.13 and 5.16 confirms the constancy of the global transfer characteristics in different reverberant laboratories and with different LSs. Therefore, the HPSC is considered independent of the measurement situation.

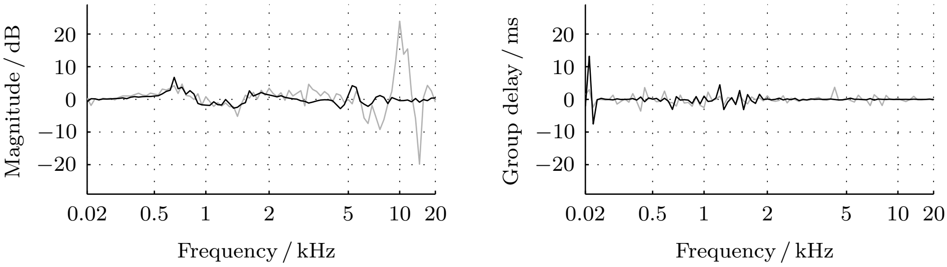

Figure 5.16: Blocked auditory canal headphone selection criterion evaluated for two different headphone specimens on an artificial head. Specimen a) is indicated by the black contours, specimen b) by the gray contours. The loudspeaker and the reproduction room are different from figure 5.13.

[65] a) Sennheiser HD 800, b) Stax λ pro NEW headphones, Klein + Hummel Studio Monitor Loudspeaker O 200, custom-made artificial head $\mathrm{AH_c}$

5.3.5 Relation to the Binaural Synthesis Quality Criterion

Comparing the HPSC defined by equation 5.39 and the overall BS error for blocked auditory canal recording given by the binaural synthesis quality criterion (BSQC) in equation 4.81, it is possible to identify contributions to the overall BS error not covered by the HPSC. The remaining error (subscript *rem*) can be formulated by

$$\mathbf{H}^{\mathrm{ah}}_{\mathrm{rem}}(\mathbf{x}_{\mathrm{m_{rec}}}, \mathbf{x}_{\mathrm{hp_{play}}}, \mathbf{x}_{\mathrm{hp_{hptf}}}, \mathbf{x}_{\mathrm{m_{hptf}}}) = \frac{\mathbf{H}^{\mathrm{ah}}_{\mathrm{hpsc}}(\mathbf{x}_{\mathrm{m_{rec}}}, \mathbf{x}_{\mathrm{hp_{play}}}, \mathbf{x}_{\mathrm{hp_{hptf}}}, \mathbf{x}_{\mathrm{m_{hptf}}})}{\mathbf{H}^{\mathrm{ah}}_{\mathrm{bsqc}}(\mathbf{x}_{\mathrm{m_{rec}}}, \mathbf{x}_{\mathrm{hp_{play}}}, \mathbf{x}_{\mathrm{hp_{hptf}}}, \mathbf{x}_{\mathrm{m_{hptf}}})}. \tag{5.43}$$

Figure 5.17 shows the transfer characteristics representing the remaining error according to equation 5.43 for the exemplary situation also described by figure 5.13. Headphone specimen a) is indicated by the black contours, specimen b) by the gray contours.

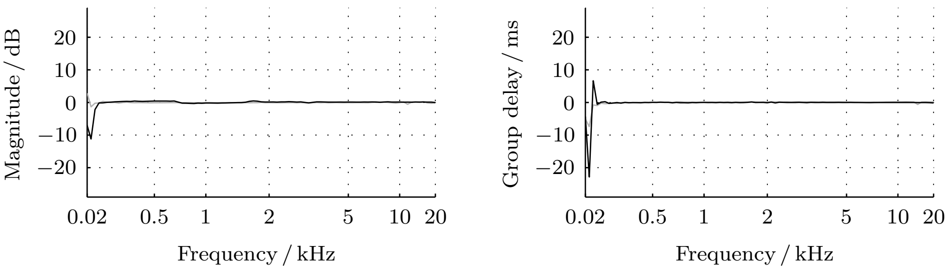

Figure 5.17: Relation of the blocked auditory canal headphone selection criterion to the blocked auditory canal binaural synthesis quality criterion. Headphone specimen a) is indicated by the black contours, specimen b) by the gray contours.

Remaining errors occur for both HP specimens in magnitude spectrum and group delay only at the lower limit of the audible frequency range. These deviations are attributed to a combination of the equalization filter design restrictions (cf. section 5.2.3) and the limited transmission bandwidth of the HPs and LSs, visible also in the BS TF shown by figure 4.8. Consequently, the errors predicted by the HPSC are confirmed by figure 5.17 to represent the overall BS error for both HP specimens.

In summary, the HPSC proposed is able to predict, based on four AH TF measurements, the suitability of HPs for BS with blocked auditory canal entrance recording and equalization based on blocked auditory canal entrance HPTFs. Speaking descriptively, the HPSC checks for the validity of assumption 7. It verifies whether the TFs connecting the sound pressure spectra at the blocked auditory canal entrances and at the eardrums, the so-called blocking factors, are identical for HP playback and for LS playback without HPs. Using an AH that enables the repeatable positioning of microphones at the eardrum positions and at the entrances of the blocked auditory canals, the procedure provides valid results in the full audible frequency range. The method proposed for predicting the suitability of HPs for BS with blocked auditory canal entrance recording by Møller (1992) provides valid results only in the frequency range below 7 kHz, primarily due to procedural shortcomings (cf. section 5.3.1 and Møller et al. 1995a).

5.4 The Auditory Canal Sound Pressure in Headphone Reproduction

For BS with HP playback, claiming identical ear signals in the reference scene and the playback situation may, contradicting assumption 2, not be sufficient to ensure identical hearing sensations: Some authors find remarkable sound pressure level differences in the auditory canals at equal loudness for LS versus HP reproduction (e. g. Fastl et al. 1985). This circumstance is according to Munson and Wiener (1952) and Rudmose (1982) referred to as *the case of the missing 6 dB*. Theile (1985, 1986) labels the suprathreshold effect discussed here *sound level loudness divergence effect (SLD-effect)* to emphasize the difference to a similar phenomenon observed near the threshold of hearing that can be attributed to physiological noise (Rudmose 1982, Theile 1985). In this chapter, the label *the case of the missing 6 dB* is used according to Rudmose (1982) since a distinction between threshold and suprathreshold effects is not required.

Based on the binaural synthesis quality criterion (BSQC) introduced by equation 4.81, figure 4.8 shows that BS, correctly implemented using conventional AH recording and conventional AH HPTFs, provides the AH reference scene TFs within the accuracy of the verification procedure. Consequently, identical sound pressure levels occur in the auditory canals for the LS reproduction of the reference scene and the HP reproduction in the playback situation. Regarding BS implemented based on blocked auditory canal entrance AH recording and HPTFs, the BSQC evaluation shown by figure 4.9 indicates influences of the HP specimens, which are also predicted by the blocked auditory canal headphone selection criterion (HPSC) introduced in section 5.3. However, HPs can be found that allow for approximately frequency independent TFs of BS with blocked auditory canal entrance AH recording and HPTFs (cf. black contours in figure 4.9). Comparable verification measurements with human heads are virtually impossible since the sound pressure distribution across the eardrum is difficult to capture (cf. section 5.1.4). However, human head verification is desired to address whether the HPSC and the BSQC are appropriate not only for the AH situation, but also for BS with human head playback.

Consequently, LTFs according to definition 5 are employed as a verification tool, in that the loudness elicited by a binaurally synthesized LS is adjusted to the loudness of the corresponding real counterpart by Békésy-Tracking (cf. section 2.5). Thereby, the loudness at approximately the same sound pressure level in the auditory canal is compared for LS and HP playback since BS by definition aims at reproducing the reference scene ear signals (cf. definition 2). This way, the ear signals are validated as the BS design goal, and aspects of individual versus nonindividual recording and HPTF measurement, discussed from a physical point of view in sections 5.1 and 5.2, are addressed perceptually.

The present section is structured as follows: After a review of the case of the missing 6 dB, the experimental setup and procedure are introduced. Subsequently, the results are presented and discussed including the verification of BS for human head playback with different equalization approaches, the justification of the ear signal recreation as a sufficient basis for BS (assumption 2), and the explanation of the missing 6 dB. Concluding, the results are summarized as a schematic working model of auditory localization and loudness perception. This section can be regarded as a verification of the discussion of HP playback in chapter 3 and of the theoretical BS aspects discussed in chapters 4 and 5.

5.4.1 The Case of the Missing 6 dB

In *Acoustic Measurements*, Leo L. Beranek constituted supra-aural HPs to require 6 to 10 dB more level at the eardrums for eliciting the same loudness as a free sound field, while the origin of this effect was largely unclear (Beranek 1949, pp. 730–731). The effect occurs for threshold measurements (Sivian and White 1933), where it is usually referred to as the difference between minimum audible field and minimum audible pressure, as well as for suprathreshold loudness adjustments. While the differences at threshold can be attributed to methodical shortcomings (Rudmose 1950, Killion 1978, Rudmose 1982), the effect at suprathreshold levels persists (Fastl et al. 1985, Keidser et al. 2000).

In 1952, Munson and Wiener presented the article *In Search of the Missing 6 Db*, reporting the effect especially at low frequencies for diotic HP presentation versus binaural listening to an LS in the free sound field. A difference between the two playback methods and therefore a possible reason for the unexpected deviation mentioned while not confirmed by Munson and Wiener was the difference of the hearing sensation positions. The effect was further observed in 1956 by Robinson and Dadson, when measuring equal loudness contours, as well as by Weingartner (1972), Theile (1984, 1985, 1986), and Stoll and Theile (1986). The latter authors found the deviations at frequencies below and above the so-called presence region around 3 kHz, for diotic HP reproduction versus LS presentation in the anechoic chamber, and to a lower extent also in the reverberant chamber.

According to Theile (1985) the effect is reduced for monaural versus binaural listening, with a free-field reference to about half the level difference and with a diffuse-field reference to almost zero, which is supported by the data of Bocker and Mrass (1959). In contrast, Goossens et al. (2009) reported the extent of the effect for third-octave band noise presented by an LS in the reverberant chamber to decay from about 6 dB for monaural comparison to about 3 dB for diotic HP presentation versus binaural listening, and to vanish for dichotic presentation of interaurally decorrelated noise or for an AH recording of the sound field created by the LS. Goossens et al. further found a correlation between the effect and the hearing sensation position in the HP condition: The effect occurs for hearing sensation positions inside the head and decreases with increasing externalization.

Possible explanations for the effect are according to Rudmose (1982) structure-borne sound transmission from the electroacoustic transducers to the subject's chair, the LS position, transducer distortions, and the procedure employed. Further, Rudmose realized that for some subjects an LS close to one ear required more level for equal loudness than a distant LS. In other words: The LS respectively the hearing sensation position may influence the adjustment results. Zollner (1995) showed that adaptation and expectation effects can influence the auditory localization process, loudness, and sound color. He further confirmed possibly missing bone conduction components during HP listening not to contribute to the effect of the missing 6 dB, as shown earlier by Genuit (1986).

Fastl et al. (1985) measured the differences $\Delta L_{\mathrm{ac}} = L_{\mathrm{hp}_{\mathrm{ac}}} - L_{\mathrm{ls}_{\mathrm{ac}}}$ between the frequency dependent auditory canal levels (subscript *ac*) of tones at equal loudness for binaural listening to diotic HP presentation versus free-field LS playback at levels around 70 dB. Figure 5.18 shows the quartiles of eight subjects' results for two different HP models. The filled circles indicate a closed HP model and the open squares an open HP model.

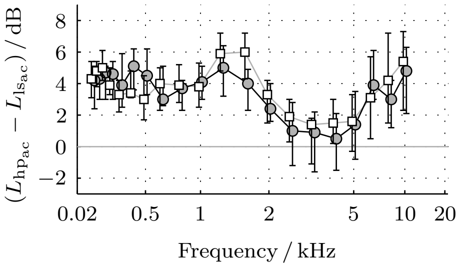

Figure 5.18: Quartiles of the level differences in the auditory canal between equally loud tones presented diotically by headphones and by a loudspeaker at 3.5 m distance for binaural listening in an anechoic chamber according to Fastl et al. (1985). Level at the listening position approximately 70 dB SPL. The filled circles indicate closed headphones, the open squares open headphones.

Qualitatively well in line with the data of figure 5.18 and quantitatively more pronounced, Keidser et al. (2000) found in the frequency range around 500 Hz, in both their own data and a literature review, on average 8 dB more level necessary to elicit equal loudness in HP versus LS playback, while Keidser et al. reported no level difference at equal loudness in the frequency range around 3 kHz. These authors also mentioned the LS position as a possible influence factor on the level difference, without validating this hypothesis.

5.4.2 Loudness Adjustment Experiment Setup

The loudness adjustment experiments presented in this section were conducted in three different rooms, two laboratories and an anechoic chamber with 250 Hz lower limiting frequency. Figure 5.19 illustrates the corresponding reverberation times.

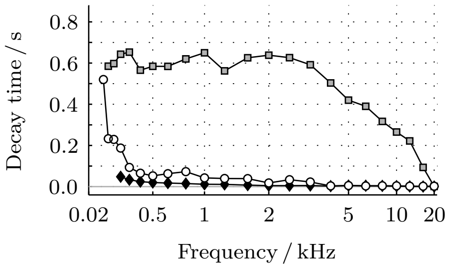

Figure 5.19: Third-octave band reverberation times (early decay times according to DIN EN ISO 3382 2000) of the rooms where the loudness adjustments presented in this section took place. Laboratory 1 (dimensions: 6.8 m × 3.9 m × 3.3 m) is indicated by white circles, laboratory 2 (6 m × 3.5 m × 3.3 m) by gray squares, and the anechoic chamber (250 Hz lower limiting frequency) by black diamonds.

As expected, the anechoic chamber (diamonds) shows almost no reverberation for frequencies above about 250 Hz. Laboratory 1 (circles) was designed to resemble a highly damped living room and therefore provides a reverberation time on average below 200 ms, globally increasing towards the lowest audible frequencies. Laboratory 2 (squares) is a typical small laboratory room with a reverberation time of about 600 ms at low and mid frequencies, decaying towards the upper limit of the audible frequency range. These acoustic conditions were included to address differences between the LTFs acquired under reverberant (laboratory 2), damped (laboratory 1), and anechoic conditions.

The static and dynamic BS systems employed were implemented, according to the theory derived in chapters 4 and 5, based on blocked auditory canal entrance recordings and HPTFs. All recordings and measurements were carried out with the same miniature microphones[66] embedded in modified foam earplugs and inserted in the auditory canals so that the canals were blocked and the microphones were positioned approximately 2 mm inside the canal. The dynamic BS system[67] was implemented with 1° grid resolution and restricted to respond to rotational head movements in the horizontal plane (unrealistic room-related binaural synthesis, cf. section 5.1.2). Therefore, a set of 360 BIRPs was used, representing one rotation of the subject with the head fixed with regard to the other body. Translational head movements induced no ear signal adaptation, in contrast to real situations. However, it is shown below that this situation simulates the ear signals occurring in a real scenario to a degree sufficient for the attempted loudness comparisons.

For the loudness adjustments, the tracking procedure introduced in section 2.5 was used in combination with pure tone stimuli. The second sound of each pair, presented by the BS, had to be adjusted to the same sound presented by the LS, which was calibrated with broadband noise to about 58 dB SPL at the listening position. Traditionally, for a perceptual verification of an HP based playback method, the subjects listen to the HPs, take them off, listen to the corresponding reference scene, indicate their judgment, and put the HPs back on before the next stimulus is presented (e. g. Zwicker and Maiwald 1963, Fastl and Zwicker 1983, Menzel et al. 2011a). This procedure is demanding for the subjects and inherently requires a temporal gap between the stimuli to be compared. To ease the procedure, BS with HP reference was used, allowing for the direct comparison of the BS situation and the reference scene by defining the reference scene for a listener wearing inactive HPs (cf. definition 30; the system-theoretic basis is discussed in appendix C). As a consequence, the binaurally synthesized and the reference scene require the listener to wear HPs, and if the same HP specimen is used in the reference scene and for playback of the BS results, the comparison must occur without taking off the HPs. Speaking descriptively, the subjects listen to the real and to the binaurally synthesized LS with the sound passing through inactive HPs. In the BS situation, the inactive (virtual) HPs are implemented by recording the BIRPs for a listener wearing the inactive HPs, while the HPs are active in the playback situation. The procedure further supports that the subjects are not provided with a non-acoustic indication on whether they are listening to an LS or the BS. This way cognitive, memory, learning, and expectation effects are attempted to be stabilized between both playback methods (cf. section 3.2.1).

In the experimental room, a chair was positioned at 1.5 m distance in front of the LS. This configuration was centered on the room midpoint, while care was taken not to position the chair or the LS directly at the midpoint. In each experimental condition, the BIRPs were recorded for a human listener seated in the chair and wearing the HPs later used to play back the BS results. For the measurements as well as the listening experiments, an optical position control based on the principle of a pinhole camera (Rayleigh 1891) was used to adjust the chair so that the midpoint of the interaural axis (cf. definition 2) of the subject seated in the chair was located horizontally and vertically with an accuracy of ±3 cm on the LS radiation axis. No head fixation was applied, and the subjects were allowed but not instructed to turn their heads freely. For the measurements, the chair was

[66] Sennheiser KE 4-211-2 electret microphones in amplifier configuration

[67] $M_\mathrm{s} = 1$, $M_\mathrm{spo} = M_\mathrm{hpo} = M_\mathrm{so} = 1$, $M_\mathrm{ho} = 360$, Klein + Hummel Studio Monitor Loudspeaker O 98

rotated around the midpoint of the interaural axis in angular steps of 6°. The intended rotational center and the step size were controlled using the head-tracking system also employed for implementing the dynamic BS[68]. Each measurement could be started only if the head position was correct with ±1 cm translational accuracy, ±0.5° horizontal rotational accuracy, and if head nodding or tilting deviations from the horizontal plane were below ±0.75°. Furthermore, after each measurement, the validity of the head position and orientation was checked again. Invalid measurements were repeated until the head position and orientation requirements were met. In a post-processing step, the horizontal BIRP grid resolution was increased to 1° by a cubic spline interpolation of the time aligned IRs independently for each side and subsequent reintroduction of the separately interpolated original delays (cf. section 5.1.2).

5.4.3 Results of the Loudness Adjustment Experiments

In this section, the LTFs from a binaurally synthesized frontal LS to the corresponding real reference scene LS acquired with eight normal hearing subjects are presented and discussed for BS systems implemented based on blocked auditory canal entrance measurements with different HPs, recording situations, and equalization methods. In this way, deviations between reference scene and recording situation (section 5.1) as well as implications of nonindividual HPTFs (section 5.2) are addressed perceptually. The order of presentation is selected with the objective of being able to discuss each situation based on the preceding data. For being acquired by listening experiments, the results must be interpreted taking into account the accuracy of the procedure (section 2.5) and the HPTF reproducibility (section 5.2). Therefore, in all figures of this section, median deviations exceeding the methodical accuracy are highlighted by gray bars on the abscissae. During the experiments, the HP specimens a) and b) were employed[69], and the respectively used specimen is indicated along with the results. Since only one specimen per model was examined, the results may not be representative of the corresponding HP model (cf. section 5.2.5).

Nonindividual Recording, Average Magnitude Equalization The dynamic BS system for the first loudness adjustment experiment is implemented according to equation 4.53 based on nonindividual blocked auditory canal entrance recording and HPTFs with average magnitude equalization, given by equation 5.23, using the filters discussed in section 5.2.4. Equation 4.53 reveals three possible error sources for this configuration: the nonindividual recording, the average magnitude equalization, and a remaining error term. Based on the HPSC given by equation 5.35, the remaining error term is attributed to unsuited HPs. For the HP specimens studied the AH HPSC is shown by figure 5.13, predicting synthesis with HP specimen b) to result in ear signals with magnitude spectra erroneous in the frequency range above some 5 kHz, while specimen a) is predicted to enable correct BS within the HPSC accuracy. Taking into account the AH BSQC shown by figure 4.9 reveals the direction predicted for the effect since the magnitude spectrum represents the

[68] Polhemus 3 Space FasTrack

[69] a) Sennheiser HD 800, b) Stax λ pro NEW headphones

frequency dependent ear signal level difference $\Delta L_\mathrm{e} = L_{\mathrm{ref}_\mathrm{e}} - L_{\mathrm{bs}_\mathrm{e}}$ expected to remain between the reference scene and the BS. The maximum ΔL_e in figure 4.9 is positive predicting a higher level in the reference scene compared to the BS, suggesting the BS provides too little energy at this frequency. Hence, to elicit the reference scene loudness, the BS is expected to require higher level driving signals than the reference scene LS in the frequency range of the ΔL_e maximum, that is between 8 and 12 kHz. Since the auditory canals of the AH used for the BSQC evaluation are smaller than average human canals (cf. figure 4.5), the characteristic frequencies are expected lower for human heads.

Initially, keeping the synthesis with nonindividual recording and average magnitude equalization constant while carrying out the experiment with different HP specimens, the HP influence is studied. Figure 5.20 shows, for the synthesis of a frontally located LS in reverberant laboratory 1 with specimen b), the quartiles of the individual LTFs, which are according to definition 5 the level differences $\Delta L_\mathrm{in} = L_{\mathrm{bs}_\mathrm{in}} - L_{\mathrm{ls}_\mathrm{in}}$ between the input signals of the binaurally synthesized and the corresponding real LS at equal loudness.

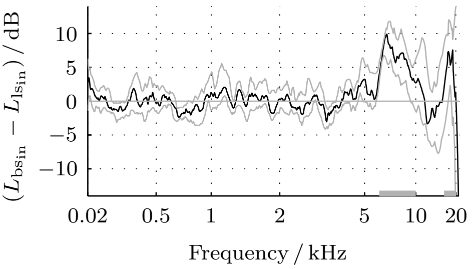

Figure 5.20: Quartiles of the individual level differences between the input signals of a binaurally synthesized and the corresponding real loudspeaker in front of the subjects in laboratory 1, adjusted to equal loudness by Békésy-Tracking. Headphone specimen b), dynamic synthesis, nonindividual recording, and average magnitude equalization. Median deviations exceeding the methodical accuracy are highlighted by gray bars on the abscissa.

The results suggest that the BS preserves the loudness of the reference scene LS well for frequencies up to some 5 kHz, as predicted for the synthesis with HP specimen b) by the AH HPSC shown by figure 5.13. In order to elicit the same loudness, the BS versus the LS input level must be increased in the frequency range between 6 and 10 kHz by about 6 dB. At frequencies between 12 and 15 kHz, the median level difference decays to about 0 dB, while at the upper limit of the audible frequency range, higher levels up to 5 dB are necessary for the BS to elicit the loudness of the LS. Based on the AH HPSC, a maximum is expected in the frequency range around 10 kHz. Therefore, it seems reasonable to assume the HP specimen contributes to the maximum between 6 and 10 kHz in figure 5.20. The lower center frequency of about 7 kHz instead of 10 kHz can be attributed to the geometric differences between the smaller AH and average human auditory canals.

Since the loudness of a fixed frequency pure tone of 400 ms duration depends solely on its level (Fastl and Zwicker 2007, pp. 205–207), frequency independent average LTFs should be achievable by adjusting the input level of the BS system according to the median of figure 5.20. Figure 5.21 shows the loudness adjustment results for the BS setup described above amended by the level correction according to the median of figure 5.20.

The resulting average LTF resembles the situations with the original BS system shown by figure 5.20. The global characteristics are somewhat more faltering, presumably due to

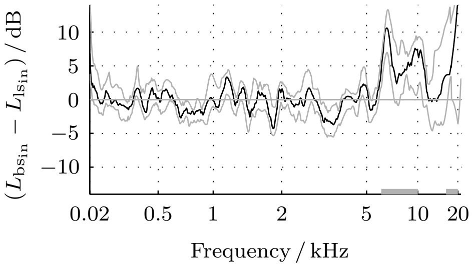

Figure 5.21: Quartiles of the individual level differences between the input signals of a binaurally synthesized and the corresponding real loudspeaker in front of the subjects in laboratory 1, adjusted to equal loudness by Békésy-Tracking. Headphone specimen b), dynamic synthesis, nonindividual recording, and average magnitude equalization. Level of the BS input signal corrected according to the black contour in figure 5.20.

the faster level changes during the tracking procedure caused by the input level correction, while the median deviations exceeding the methodical accuracy remain. Therefore, the equalization of the system by an input level correction is *not* possible. This fact suggests resonances in the HP-ear system to be responsible for the level differences by causing destructive interference effects (cf. section 5.3).

As predicted by the HPSC, dynamic BS with blocked auditory canal entrance miniature microphone recording, average magnitude equalization based on blocked auditory canal HPTFs, and HP specimen b) causes undesired effects on the loudness transfer. These effects are not compensable by BS input level correction. The deviations are qualitatively in line with the AH approximation of the HPSC, indicated for specimen b) by the gray contours in figure 5.13.

Based on the HPSC, frequency independent LTFs of BS with blocked auditory canal recording are more likely to result with HP specimen a), indicated by the black contour in figure 5.13. To validate this prediction, figure 5.22 shows the LTFs for dynamic BS of the reverberant laboratory room 2, implemented using HP specimen a) with nonindividual blocked auditory canal entrance recording following equation 4.53 and average magnitude equalization according to equation 5.23 based on blocked auditory canal entrance HPTFs.

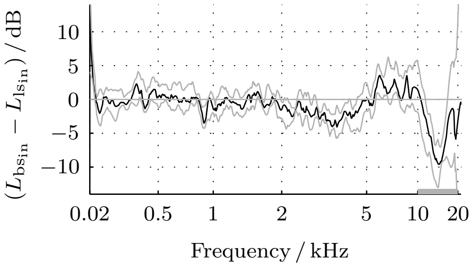

Figure 5.22: Quartiles of the individual level differences between the input signals of a binaurally synthesized and the corresponding real loudspeaker in front of the subjects in laboratory 2, adjusted to equal loudness by Békésy-Tracking. Headphone specimen a), dynamic synthesis, nonindividual recording, and average magnitude equalization.

The results shown by figure 5.22 suggest that the BS is capable of approximately reproducing the loudness of the real LS in the frequency range below some 10 kHz, apart from a reduced loudness of the BS at frequencies below about 100 Hz. This low-frequency effect is attributed to a combination of constraints regarding the nonindividual equalization filter design (cf. section 5.2.3) and the limited transmission bandwidths of the HPs and the LS, as discussed regarding the BS TF in the context of figure 4.8. The increased BS level in the frequency range from 6 to 10 kHz visible in figure 5.20 is reduced, whereas between 10 and 20 kHz, the BS input level is adjusted lower than expected.

Since an influence of the reproduction room on the loudness adjustment is possible, the experiment was repeated with the same configuration in an anechoic chamber (cf. figure 5.19). The corresponding results are shown by figure 5.23.

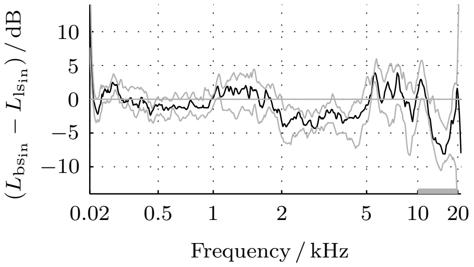

Figure 5.23: Quartiles of the individual level differences between the input signals of a binaurally synthesized and the corresponding real loudspeaker in front of the subjects in the anechoic chamber, adjusted to equal loudness by Békésy-Tracking. Headphone specimen a), dynamic synthesis, nonindividual recording, and average magnitude equalization.

The most prominent differences of figure 5.23 versus figure 5.22 are the globally enlarged inter-quartile range and the higher median level in the frequency range between 1 and 2 kHz, which is the starting region of the tracking procedure. The enlarged inter-quartile range may be due to procedural reasons since all subjects perceived the adjustment procedure in the anechoic chamber more demanding and the listening situation less familiar compared to the reverberant laboratories. As a consequence, the adjustments required more time under anechoic conditions, resulting in a broader frequency range for the initial calibration of the starting level to the level at equal loudness. However, the originally addressed deviations in the frequency range between 10 and 20 kHz remained comparable in the anechoic chamber (figure 5.23) and laboratory 2 (figure 5.22). In summary, the acoustical environments emphasize methodical artifacts differently, but show no further characteristic influences on the LTF of BS.

Two possible causes for the high frequency artifact remain: the average magnitude equalization and the nonindividual recording. To address the equalization influence, the discussion proceeds with nonindividual recording and nonindividual versus individual magnitude equalization. Finally, after showing that static BS is able to approximate the loudness transfer of dynamic BS, individual recording is evaluated with average magnitude, individual magnitude, and individual magnitude and phase equalization using static BS.

Nonindividual Recording, Nonindividual Equalization This paragraph addresses average versus nonindividual magnitude equalization in dynamic BS with nonindividual blocked auditory canal entrance recording according to equation 4.53. The experiment in the anechoic chamber represented by figure 5.23 was repeated with nonindividual magnitude equalization according to equation 5.22 using blocked auditory canal entrance HPTFs of the subject employed for the BIRP recording. The remaining configuration was not modified; this and all further experiments were carried out with HP specimen a). Figure 5.24 shows the results.

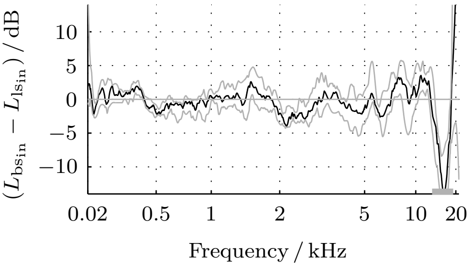

Figure 5.24: Quartiles of the individual level differences between the input signals of a binaurally synthesized and the corresponding real loudspeaker in front of the subjects in the anechoic chamber, adjusted to equal loudness by Békésy-Tracking. Headphone specimen a), dynamic synthesis, nonindividual recording, and nonindividual magnitude equalization.

Figure 5.24 reveals that with nonindividual recording and the corresponding nonindividual magnitude equalization the high-frequency artifact remains visible but becomes narrower compared to the average magnitude equalization situation depicted by figure 5.23. Therefore, in combination with nonindividual recording, neither average nor nonindividual magnitude equalization allow for frequency independent LTFs. This result is to be expected based on the inter-individual HPTF variability characteristics shown by figure 5.11, which indicate inter-individually different magnitude spectra without equalization as well as with average magnitude equalization. Therefore, individual magnitude equalization is addressed in the next paragraph.

Nonindividual Recording, Individual versus Average Magnitude Equalization Since the listening environments did not influence the characteristic structure of the LTFs (cf. figure 5.22 versus figure 5.23) and since the procedure was considered less demanding by the subjects under reverberant conditions, this and all further experiments were conducted in reverberant laboratory 2. Figure 5.25 shows, for dynamic BS with HP specimen a) and nonindividual blocked auditory canal entrance recording as given by equation 4.53, the comparison of the inter-individual medians resulting from the loudness adjustments with average (gray contour, equation 5.23) and individual magnitude equalization (black contour equation 5.22), based on blocked auditory canal HPTFs.

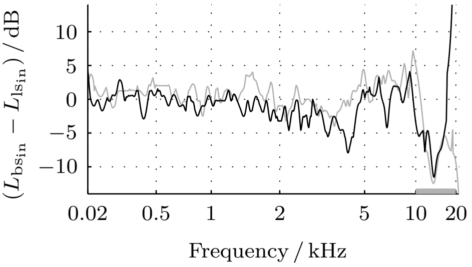

Figure 5.25: Median of the individual level differences between the input signals of a binaurally synthesized and the corresponding real loudspeaker in front of the subjects in laboratory 2, adjusted to equal loudness by Békésy-Tracking. Headphone specimen a), dynamic synthesis, nonindividual recording, and individual (black contour) versus average magnitude equalization (gray contour).

According to figure 5.25, similar results are obtained for average and individual blocked auditory canal equalization. Therefore, nonindividual blocked auditory canal entrance recording does not enable frequency independent LTFs. This result is presumably due to a mismatch in the high frequency characteristics of the nonindividual BIRPs, the individual equalization filters, and the actual human heads in the playback situation. In order to verify this hypothesis, individual recording is addressed in the following. However, due to the effort of recording individual BIRPs for a full rotation of the subject, static instead of dynamic BS (cf. definition 24, p. 52) would be preferable for the loudness adjustment experiments. That way, the effort could be reduced for the BS system used to one instead of sixty recordings per ear. To justify employing the static procedure, a comparison between static and dynamic nonindividual BS with average magnitude equalization is given in the next paragraph, before individual blocked auditory canal recording is discussed.

Dynamic versus Static Nonindividual Recording, Average Magnitude Equalization Figure 5.26 shows the inter-individual medians of loudness comparison results acquired with static BS, indicated by the black contour, and medians resulting from the same experiment with dynamic BS, represented by the gray contour. Both BS systems were implemented identically, apart from the static or dynamic property, with HP specimen a), simulating reverberant laboratory 2 using nonindividual blocked auditory canal entrance recording according to equation 4.53 and average magnitude equalization, implemented following equation 5.23 based on blocked auditory canal HPTFs.

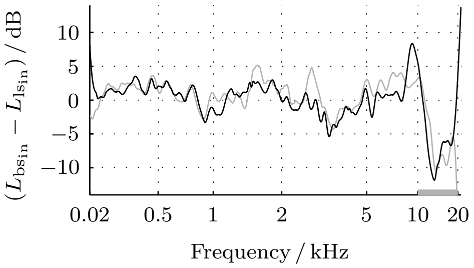

Figure 5.26: Median of the individual level differences between the input signals of a binaurally synthesized and the corresponding real loudspeaker in front of the subjects in laboratory 2, adjusted to equal loudness by Békésy-Tracking. Headphone specimen a), static (black contour) versus dynamic (gray contour) synthesis, nonindividual recording, and average magnitude equalization.

Based on the results shown by figure 5.26, static BS is assumed to resemble dynamic BS regarding the loudness transfer studied in this section with an average deviation smaller than the average accuracy of the LTF measurement method employed, which is given in section 5.4.2 to ±2 dB. All subjects reported rather similar externalized hearing sensation positions elicited by the real LS and its binaurally synthesized counterpart, for static as well as dynamic BS. Consequently, the LTFs acquired with the static BS are considered representative of the respective dynamic BS system in the situation studied here, and the experiments discussed in the following paragraphs were conducted with static BS.

Individual Recording, Average Magnitude Equalization In the present paragraph, individual recording in combination with average magnitude equalization is discussed. Figure 5.27 shows the loudness comparison results for static BS with individual blocked auditory canal entrance recording as given by equation 4.53 and average magnitude equalization, implemented following equation 5.23 based on blocked auditory canal entrance HPTFs. The loudness adjustment experiment was conducted in the reverberant laboratory 2 using HP specimen a).

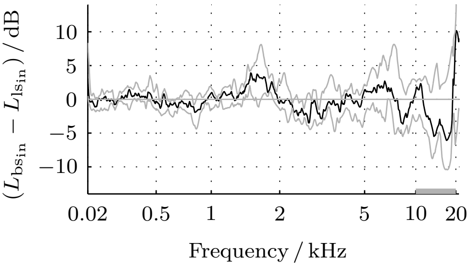

Figure 5.27: Quartiles of the individual level differences between the input signals of a binaurally synthesized and the corresponding real loudspeaker in front of the subjects in laboratory 2, adjusted to equal loudness by Békésy-Tracking. Headphone specimen a), static synthesis, individual recording, and average magnitude equalization.

According to figure 5.27, less input level is necessary for the binaurally synthesized LS compared to the reference scene LS in the frequency range between 10 and 20 kHz to elicit the reference scene loudness also for individual recording with average magnitude equalization. However, the extent of the effect is reduced compared to the situations with nonindividual recording using the same HP specimen (shown by figures 5.22 to 5.26).

Individual Recording, Individual Equalization Individual BS equalization may either aim at equalizing the magnitude spectrum only, for example using a filter with linear phase characteristics according to equation 5.22, or attempt to correct the magnitude spectrum and phase characteristics of the non-equalized BS, as formulated by equation 4.54. Both methods are discussed in this paragraph.

Figure 5.28 shows the loudness adjustment results for static BS with individual blocked auditory canal entrance recording, as given by equation 4.53, and individual magnitude equalization, implemented according to equation 5.22 based on HPTFs measured at the blocked auditory canal entrance. The experiment and the BIRP recording took place in reverberant laboratory 2 using HP specimen a).

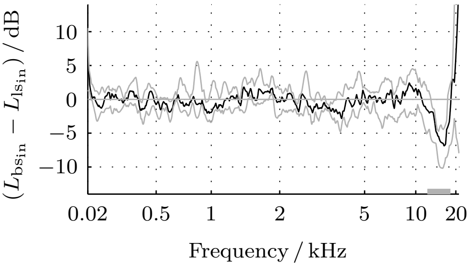

Figure 5.28: Quartiles of the individual level differences between the input signals of a binaurally synthesized and the corresponding real loudspeaker in front of the subjects in laboratory 2, adjusted to equal loudness by Békésy-Tracking. Headphone specimen a), static synthesis, individual recording, and individual magnitude equalization.

In the frequency range below 12 kHz, the LTFs depicted by figure 5.28 are frequency independent within the accuracy of the procedure (cf. section 5.4.2). However, when combining individual magnitude equalization with individual recording, a deviation exceeding the methodical accuracy occurs in the frequency range between 12 and 18 kHz. Therefore, figure 5.29 shows the inter-individual quartiles of the LTFs for the same configuration, based on blocked auditory canal entrance HPTF measurement and BIRP recording, but with individual magnitude *and* phase equalization according to equation 4.53.

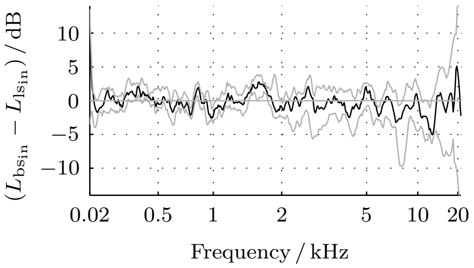

Figure 5.29: Quartiles of the individual level differences between the input signals of a binaurally synthesized and the corresponding real loudspeaker in front of the subjects in laboratory 2, adjusted to equal loudness by Békésy-Tracking. Headphone specimen a), static synthesis, individual recording, and individual magnitude and phase equalization.

On average, the results represented by figure 5.29 are considered frequency independent within the accuracy given by the adjustment procedure employed. The enlarged inter-quartile range at frequencies above some 12 kHz, compared to the situation with magnitude equalization only (figure 5.28), may result from the combination of the phase equalization with the static synthesis and the intra-individual group delay variability of the HPTFs, which reaches according to figure 5.5 a) considerable values especially at high frequencies. Due to the intra-individual variability of the high-frequency HPTF group delay characteristics caused by varying HP positions, static synthesis and reproducibility issues become more critical if the BS phase equalization is attempted (cf. section 5.2.1).

The results shown by figure 5.29 reveal that individually equalized static BS with individual blocked auditory canal entrance recording, as given by equation 4.53, is on average capable of recreating the reference scene loudness over the full audible frequency range, at least with the accuracy of the loudness adjustment procedure employed (±4 dB, cf. section 2.5). The results further support the validity of equation 4.43 and the HPSC shown by figure 5.13, which predict approximately no deviation between the BS situation and the corresponding reference scene for the configuration represented by figure 5.29.

5.4.4 Locating the Missing 6 dB

Fastl et al. (1985) stated that *"typically the sound signals at the eardrums are regarded as the most essential acoustical input parameters leading to auditory sensations in subjects* [; . . .] *on the contrary* [. . .] *tones from loudspeaker versus headphones can be perceived with different loudness despite equal sound level in the auditory canal."* However, the situations contrasted here are indeed not contradictory since the *sound signals at the eardrums* are time dependent signals and therefore not fully described by the sound level based on the root-mean-square value of sound pressure or sound intensity. Especially interaural phase relations are not reflected by the sound level.

As confirmed by the preceding section, it is accurate to state that *tones at equal level in the auditory canal* can be *perceived with different loudness.* The claim of identical ear signal *time functions* in HP and LS reproduction on the contrary actually results in equal loudness at equal auditory canal level. This is proven by the approximately frequency independent average LTF for BS with individual blocked auditory canal recording and equalization (figure 5.29) since BS aims at recreating the reference scene ear signals, the sound pressure *time signals* at the eardrums. Comparing the LTFs for BS based on individual recording with individual magnitude equalization (figure 5.28) versus with individual magnitude and phase equalization (figure 5.29) reveals an influence of the phase equalization on the loudness transfer even for the pure tone stimuli applied.

The data may be interpreted as follows: Comparing the HP and LS situations with the same root-mean-square sound pressure levels in the auditory canals does not assure the same ear signals, that is the time dependent sound pressure signals at the eardrums, in both situations. The actual sound pressure time function, not only its envelope or average, can influence auditory perceptions, reflected for example in masking-period patterns (Zwicker 1976a,b,c,d, Fastl and Zwicker 2007, pp. 93–97), binaural masking level differences (e. g. Zwicker and Henning 1991), and variations of the sound color or the hearing sensation positions (Genuit 1986). It is apparently not in every case sufficient to ensure the reference scene sound levels in the auditory canals for eliciting the reference scene loudness (e. g. Fastl 1986). Considering the results of the loudness adjustment experiments, it may be concluded that the accuracy of the ear signals achieved with individually equalized static BS, implemented using individual blocked auditory canal entrance recording, is sufficient for eliciting on average equal loudness perceptions with HP and LS presentation, for seated subjects facing the sound source, without the sophisticated methods and procedures described by Rudmose (1982).

Based on the results presented in this section (cf. Völk et al. 2011d, Völk and Fastl 2011a), the case of the missing 6 dB can be explained: Different auditory canal levels in LS and HP reproduction may arise at equal loudness since the same sound pressure time functions at the eardrums, at least to the degree achieved by the static individual BS discussed, are necessary to ensure equal loudness. The results further show that the recreation of the reference scene ear signals can be considered the ultimate goal of BS, justifying assumption 2. Thereby, instrumental free-field equalization (section 3.2.4, Zwicker and Maiwald 1963, Villchur 1969) is shown to be possible by correctly implemented static or dynamic BS with BIRPs representing anechoic conditions. BS with BIRPs recorded in a reverberant chamber allows for instrumental diffuse-field equalization (Theile 1986).

5.4.5 Implications on Psychoacoustic Modeling

Auditory impressions in general include the positions of the hearing sensations. The hearing sensations are typically located differently for single LS reproduction, where they normally lie close to or at the LS position (localization), and for conventional HP listening, where the hearing sensations usually arise inside the listener's head (lateralization, cf. section 3.1.1, Jeffress and Taylor 1961, Plenge 1974). Theile (1980) proposed the so-called *association principle*, modeling the hearing system after the peripheral processing of the inner ear (cf. section 2.4) by a pattern recognition system decoding the hearing sensation positions before addressing other hearing sensation properties. The peripheral processing is frequently simulated by adaptive filter-bank approaches (Zwicker 1979, 1986, Jin et al. 2000, Jepsen et al. 2008), but also mechanical models exist (e. g. Epp et al. 2010). Theile (1980) described the position decoding as an adaptive filtering procedure, removing the position information from the peripherally processed signals during the localization process. The resulting signals without position information are passed on to the subsequent processing stages (Theile 1981). Position information is according to Theile (1984) encoded by the direction dependent outer ear transfer characteristics, described system theoretically by the BIRPs defined in section 4.2 and discussed in detail by Blauert (1997, pp. 78–93). From an engineering point of view, this approach appears meaningful regarding the hearing process as a communication and information acquisition procedure, for each source of interest primarily attempting to extract the information contained in the source signal, while separately estimating the possibly time-variant transfer paths. From the transfer path characteristics, source location information and control factors can be extracted, which are then accessible to further processing stages such as noise reduction, beam-forming, or adaptive signal processing in general, for example with the aim of auditory scene analysis (e. g. Schwartz and Shinn-Cunningham 2010).

Schematic Working Model of the Hearing Sensation Buildup In this paragraph, the filter adaptation dependent on the sound incidence direction of Theile's model is extended to cover situations where the localization process and therefore the optimal filter adaptation and directional information removal fail. Further, a refined discussion of the adaptation process and the related auditory processing is given, with the aim of accounting for the data discussed in this thesis. The schematic working model proposed in the following aims to summarize the results derived here, not to establish a comprehensive and all-embracing independent model of the human auditory system. Indicating the descriptive nature of the discussion of physiological respectively neurological mechanisms, the terms *filtering* and *signal* are used within this section in a descriptive, not in a system-theoretically defined manner, in contrast to the nomenclature employed elsewhere in this work.

Cognitive, memory, and adaptation effects as well as inter-modal interactions are considered external control factors in the schematic working model. However, the importance of these control factors is stressed, as for example visual stimulation is known to possibly influence or even dominate the auditory perception (Völk et al. 2010b, Menzel 2011). Also adaptation effects can modify hearing sensations (Zollner 1995). Furthermore, Hubel et al. (1959) showed that attention, and therefore the brain state, may influence the way stimuli are processed and perceived. These examples illustrate the complex nature of the hearing sensation buildup, which is not modeled in full detail here. However, the examples also indicate the necessity of time variant, partially signal and system state dependent and therefore temporally extended forward and backward interaction mechanisms between all stages of hearing system models (cf. Blauert et al. 2009).

Following Theile (1980), the auditory localization process is modeled by a pattern matching system, supplemented with a source recognition and segregation module. The spectrally subdivided peripheral processing results are referred to as filter-bank output signals here, regardless of the actual implementation. For the pattern matching, hearing sensation positions are represented internally by temporally and spectrally extended basis functions, which contain the information encoded by the corresponding BIRPs. The system aims for each source recognized by the source recognition module at fitting the basis functions to the source wise segregated filter-bank output signals. Per recognized source, the segregation is assumed to produce one set of filter-bank output signals. Binaural mechanisms and adaptations induced by control factors (e. g. cognition, memory, and other modalities) are assumed to occur at each system stage, including the peripheral actively controlled basilar membrane. The overall localization process is regarded as a continuous, inter-aurally related spectro-temporal comparison of all available basis functions to the sets of filter-bank output signals. Per recognized source, correlation coefficients are continuously computed for all basis functions. The correlation coefficients indicate the similarity of the basis functions, representing hearing sensation positions, and the sets of filter-bank output signals, representing sound sources. Additionally, the correlation coefficients can be modified by the control factors, which for example represent externally induced adaptations or, combined with a memory stage, temporally extended mechanisms as for example the direct-to-reflected sound energy ratio detection (Bronkhorst and Houtgast 1999) or the precedence effect buildup (Wallach et al. 1949, Houtgast and Aoki 1994). For each recognized source, the basis functions weighted by the modified correlation coefficients are combined to the transfer characteristics compensated by the adaptive filters before the signal set representing the source is passed on to the subsequent processing stage. Accordingly, the hearing sensation positions elicited by each source are determined based on the locations represented by the basis functions, taking into account the modified correlation coefficients by a statistical spatial decision process.

Incorporating localization experiment results for unnatural and therefore presumably hard to classify stimuli (Blauert 1997, pp. 137–177), the hearing sensation position arising if insufficient correlation is found is assumed inside the head. Speaking descriptively, lateralization (in-the-head localization, cf. section 3.1.1) occurs based on the working model if the localization process for the respective source fails. In this case, the filtering effect is reduced by small correlation coefficients, and the virtually unprocessed signal set representing the lateralized source is passed on to later system stages. Consequently, erroneously remaining signal components can influence loudness, sound color, and other hearing sensation properties in a possibly sound source position dependent way. Relations between loudness and hearing sensation respectively sound source positions have been observed by different authors (cf. section 5.4.1, Munson and Wiener 1952, Rudmose 1982, Keidser et al. 2000, Völk et al. 2011d, Völk and Fastl 2011a).

Regarding the amount of correlation required for hearing sensations to be externalized, which is to appear outside the head, no strict limit is assumed. Distance perception experiments with suboptimal BS systems (Völk et al. 2008a) show gradually decreasing externalization when modifying BS from using individual, to nonindividual, to AH BIRPs (cf. section 5.1.3). The decreasing degree of individualization results in less correlation detected by the pattern matching of the working model. This is supported by the results of Goossens et al. (2009), who reported the effect of the missing 6 dB to gradually diminish with increasing externalization in the HP condition. Underestimating the distance of a presumably incorrectly localized source, possibly representing a threat or danger, is further plausible from an evolutionary point of view.

From an engineering perspective, it appears reasonable to design the adaptive filters removing successfully matched basis functions not to attenuate signals modified by the filters versus signals modified less or not at all. For all sound incidence directions, the outer ear transfer characteristics are dominated by the first auditory canal resonance in the frequency range around 3 kHz (Blauert 1997, pp. 86–93). If the filters intended to compensate for the outer ear transfer characteristics are designed to amplify spectral components outside the resonance frequency range, rather than attenuating the resonance, a relative amplification of the sets affected by the filters occurs compared to less or not affected sets. In other words, sets of filter-bank output signals corresponding to successfully localized sources are amplified outside the resonance frequency range compared to sets representing sources without conclusive localization results. This procedure is assumed for the working model, which appears meaningful considering communication in noise, where diffuse noise with distributed sources is unlikely producing a localization result, whereas the communication partner's voice is amplified with respect to the background noise if the localization process succeeds (cocktail party effect, cf. Bodden 1993, Roman et al. 2003). The resulting relative attenuation of low-frequency noise components is especially preferable taking into account the level dependent upward-spread of masking (cf. Fastl and Zwicker 2007, pp. 64–74).

Application to Earlier Experimental Results Figure 5.18 shows the *auditory canal level difference of equally loud tones* for binaural listening to diotic HP versus LS reproduction in an anechoic chamber according to Fastl et al. (1985). Applying the working model proposed above, the data may be interpreted as follows: Assuming the localization process succeeds when listening to the LS, the adaptive filters correct for the outer ear transfer characteristics for frontal sound incidence, which are reported for example by Blauert (1997, p. 86). Since the level maximum in the frequency range of the first auditory canal resonance at about 3 kHz is assumed as the internal reference, components at lower and higher frequencies are amplified with regard to the 3 kHz region before the filter-bank output signal set is passed on to the subsequent processing stages, which then form loudness and other hearing sensation properties. Further assuming inside-the-head localization when listening to the HPs, the corresponding filter-bank output signals also include the auditory canal resonance but are not affected by the adaptive filters before being passed on to the loudness evaluation. Consequently, for eliciting equal loudness, the HPs must be driven by more level in the spectral regions where the adaptive filters are effective for LS listening. Since both sets of filter-bank output signals contain the first auditory canal resonance and are not modified by the adaptive filters in the resonance frequency range, no level difference at equal loudness occurs in this frequency range. In summary, the working model accounts qualitatively for the results of Fastl et al. (1985).

According to Theile (1985), the auditory canal level differences at equal loudness for HP versus LS reproduction in an anechoic chamber are reduced for *monaural versus binaural listening* (cf. definition 3), which is supported by Bocker and Mrass (1959). Apparently, these results contradict the data of Goossens et al. (2009), who report larger level differences at equal loudness for monaural versus binaural comparisons of HP and LS reproduction in a reverberant chamber. The findings of Theile (1985) and Bocker and Mrass (1959) can be explained by the working model assuming the monaural LS localization process results in less correctly matched basis functions and therefore provides lower overall correlation than the binaural localization process. As a consequence, less adaptive filtering occurs in the monaural compared to the binaural LS listening situation, resulting in less auditory canal level difference to diotic HP reproduction at equal loudness. In the situation discussed by Theile (1985), also the HP reproduction is modified from diotic to monotic presentation (cf. definition 13). Monotic HP playback can elicit a hearing sensation localized close to or at the active HP capsule (Blauert 1997, p. 158). In this case, the working model predicts enlarged correlation factors and therefore more adaptive filtering than for a lateralized hearing sensation with diotic HP playback. This effect further reduces the auditory canal level difference at equal loudness, as the adaptive filter settings in the HP and LS situations become increasingly similar. Regarding the results of Goossens et al. (2009), the hearing sensation positions are expected close to or within the head for diotic HP presentation and not clearly localized for binaural listening to an LS in the reverberant chamber (Blauert 1997, pp. 158 and 242). The working model suggests little correlation in both conditions and therefore little auditory canal level difference at equal loudness. This is supported by the results in the binaural conditions of Goossens et al. (2009) and of Theile (1985). Furthermore, hardware configurations providing a higher overall correlation for the monaural versus the binaural LS localization process in the reverberant chamber may be possible. If in this case the monotic HP presentation elicits a lateralized hearing sensation, the working model also accounts for the results of Goossens et al. (2009) in the monaural condition, which show larger level differences between HP and LS presentation at equal loudness than the binaural comparison.

The phenomenon typically referred to as *directional bands* denotes illusory hearing sensation positions elicited by amplifying specific frequency ranges of the ear signals (Blauert 1997, pp. 108–115). The working model applied to this effect suggests that the physical ear signal modifications cause increasing correlation between the basis functions and the set of filter-bank output signals representing the source under consideration. If the physical sound variations correspond sufficiently to the outer ear TFs representing the intended position, the corresponding location is perceived and the correctly matched basis functions are removed from the signal set passed on to the subsequent processing stages. Contributions of the physically applied signal modifications not covered by the matched basis functions are handed on to the subsequent processing stages and can explain the unintended sound coloration possibly associated with the intended localization effect.

The *loudness adjustment results* of this chapter show that the loudness elicited by BS converges to the reference scene loudness as, with increasing individualization, the ear signals gain authenticity. This is predicted by the working model since with growing BS individualization, an increasing number of basis functions is matched correctly, resulting in raising overall correlation and more correct localization. As the correlation becomes identical to the reference scene, the signals passed on to the loudness evaluation process in the BS situation and the reference scene become identical, resulting in equal loudness.

It may be concluded from the working model that equal loudness (and identical hearing sensations) in LS and HP reproduction at the same auditory canal level can only occur if the hearing sensation positions are identical. Expressed more generally: Sound color respectively loudness and auditory localization are related, which becomes especially apparent by the coloration artifacts occurring if the localization process fails.

5.5 Summary

The binaural synthesis theory is derived based on several assumptions in chapter 4. In the present chapter, the validity of these assumptions is discussed, along with physical and perceptual consequences arising if the assumptions are violated. Thereby, the authenticity of hearing sensations achievable with binaural synthesis implemented with blocked auditory canal entrance recording and headphone transfer functions is evaluated.

In the initial section, deviations between reference scene and recording situation are addressed, including theoretical and practical aspects of dynamic binaural synthesis, resulting in a perceptually motivated procedure for the determination of grid resolution and signal processing requirements. The results show that the effort required for transparent binaural synthesis depends on the situation to be simulated. In general, the most authentic systems are implemented based on individual binaural impulse response pairs. Reducing the effort by employing nonindividual human head recording or typical currently available artificial heads gradually degrades the authenticity of the resulting hearing sensations. For the different recording situations, the inter-individual variability is addressed quantitatively, and systematic as well as procedural aspects are discussed. Regarding inversion problems and dynamic range limitations in the implementation of binaural synthesis systems, auditory-adapted exponential transfer function smoothing (AAS) is proposed and evaluated, a procedure aiming at the reduction of spectral fluctuation while not diminishing the amount of relevant auditory information.

The second section covers deviations between the headphone transfer function measurement situation and the reference scene as well as the recording situation. For the circum-aural headphone specimens studied, the reproducibility characteristics of the headphone transfer functions and microphone transfer functions show comparable frequency dependencies: Magnitude spectrum and group delay variability values due to repositioning increase with frequency, reaching auditory relevant amounts in the range above about 6 kHz, with average values of ±2 dB and ±100 μs at the upper limit of the audible frequency range. Local variability maxima arise in the frequency range of dips in the magnitude spectra. Regarding the inter-individual and inter-specimen differences of headphone transfer functions, for the same circum-aural specimens a variability structurally comparable to the intra-individual variability due to the headphone repositioning occurs, but with about ±3 dB and ±50 μs increased magnitude.

The blocked auditory canal headphone selection criterion (HPSC) for binaural synthesis is proposed in the third section. This criterion predicts, based on four artificial head transfer function measurements, the suitability of headphones for binaural synthesis implemented based on blocked auditory canal entrance recording and headphone transfer function measurement. Speaking descriptively, the HPSC verifies whether the transfer functions connecting the sound pressure spectra detected by miniature microphones at the entrances to the blocked auditory canals and the sound pressure spectra at the eardrums, referred to as blocking factors, are identical for headphone playback and loudspeaker reproduction without headphones. Using an artificial head allowing for the repeatable positioning of the microphones at the eardrum locations and at the entrances to the blocked auditory canals, the procedure provides valid results in the full audible frequency range. The method proposed for the same purpose by Møller (1992) allows for valid results only at frequencies below about 7 kHz, primarily due to procedural reasons. Further, Møller's method requires probe microphones, which provide less signal to noise ratio than the miniature and artificial head microphones used to acquire the HPSC.

Summarizing the fourth section, taking into account different binaural synthesis implementations, the loudness elicited by narrow-band signals is compared for presentation over a binaurally synthesized virtual loudspeaker and the corresponding real counterpart. The results indicate that the loudness elicited by the loudspeaker can be approximated over the full audible frequency range. This holds true for static binaural synthesis with individual recording combined with individual magnitude and phase equalization, implemented based on blocked auditory canal entrance measurements with headphones fulfilling the HPSC proposed in section 5.3. Thereby, the HPSC is confirmed by loudness adjustment results. Less elaborate binaural synthesis procedures are capable of recreating the loudness evoked by the loudspeaker at frequencies below about 6 kHz, while eliciting deviating perceptions at higher frequencies. The deviations can be attributed to nonindividual measurement (comparing figures 5.22 and 5.27), nonindividual equalization (cf. figures 5.27, and 5.29), and inappropriate headphones (cf. figures 5.20 and 5.22 and section 5.3). As a consequence, reproducing the reference scene ear signals is shown to be the valid binaural synthesis design goal. The negligible difference between static and dynamic binaural synthesis shown by figure 5.26 reveals no considerable influence of dynamic ear signal variations due to head movements on the loudness perception in the rather static binaural synthesis reference scene discussed in the present section. Further, an explanation of the case of the missing 6 dB is derived from the results: Identical sound pressure time functions detected by the eardrums ensure equal loudness in loudspeaker and headphone reproduction, which is not necessarily given for equal auditory canal levels. Finally, a descriptive working model of auditory localization and loudness extending Theile's association principle is deduced to summarize the results, suggesting interrelations of loudness respectively sound color perception and auditory localization.

The discussion of binaural synthesis is initiated at the beginning of chapter 4 by a quote of Møller (1992), formulating the basic idea of binaural recording: By reproducing the ear signals *"the complete auditive experience is assumed to be reproduced, including timbre and spatial aspects."* Møller especially emphasized the hearing sensation properties timbre and spatial aspects, which points out that he considered these properties important indicators of the quality of binaural synthesis systems. This importance is confirmed by the considerations of Genuit (1986) and Wightman and Kistler (2005). The working model proposed in this thesis suggests that the above-mentioned hearing sensation properties are even interrelated: A general and situation independent accurate timbre respectively sound color reproduction requires correctly set adaptive filters and therefore a successful localization process. According to the working model, *"the complete auditive experience"* includes *"timbre and spatial aspects"* not only literally but also for procedural reasons.

6 Conclusions

Virtual acoustics procedures aim at recreating hearing sensations of a real or hypothetical reference scene. *Hearing research* primarily addresses the functionality of the human hearing system. Both fields of research are *interrelated* regarding different aspects: The methods of virtual acoustics may support hearing research in eliciting specific hearing sensations or by providing defined sound stimuli, as for example required in Psychoacoustics. In turn, results of hearing research can support the development of optimized virtual acoustics procedures, and the methods of hearing research, especially of Psychoacoustics, may be used for the auditory evaluation of the quality of virtual acoustics systems.

This thesis provides a *theoretical and methodical framework* for employing virtual acoustics systems for the audio playback in hearing research as well as for the auditory quality evaluation of virtual acoustics systems using methods of Psychoacoustics. The framework is verified by the headphone based virtual acoustics procedure *binaural synthesis*. Consequently, this work further provides a system-theoretic derivation of static and dynamic binaural synthesis, including a physical and psychoacoustical evaluation of the achievable synthesis results, especially with regard to the application of binaural synthesis as the audio-playback procedure in hearing research.

Primary methodical contributions of this work are refined formulae for the frequency dependence of critical bandwidth and critical-band rate, as well as accordingly implemented tools for the perceptual and instrumental analysis of audio signals and transmission systems. The closed *mathematical formulation of the critical-band concept* is designed to allow for the direct parameterization of auditory-adapted algorithms. In contrast to earlier analytical expressions, the formulae reflect the critical-band concept in the full audible frequency range by an invertible critical-band rate function and a critical bandwidth formula converging to $\Delta f = 0\,\mathrm{Hz}$ at $f = 0\,\mathrm{Hz}$, with $\Delta f \leq 2f \;\forall\, f$.

Using the revised formulation of the critical-band concept, a combined signal and system analysis procedure, referred to as *auditory-adapted analysis* (AAA), is derived. AAA models peripheral auditory mechanisms aiming at visualizing auditory relevant system or signal characteristics in a technically conclusive way comparable to established visualizations. Audio signal processing systems are therefore classified regarding their audible impact as either spectro-temporally or purely spectrally effective. For the analysis of spectro-temporally effective systems or audio signals, an auditory-adapted spectrogram representation is introduced, visualizing magnitude and phase information by a single image. Purely spectrally effective systems on the contrary are analyzed based on their auditory-adapted transfer characteristics, displaying the frequency dependent magnitude and group delay computed from an auditory-adapted Fourier spectrum.

The proposed critical-band formulations further serve as the conceptual basis for a procedure referred to as *auditory-adapted exponential transfer function smoothing* (AAS) which attempts inaudible spectral transfer function smoothing. AAS is proposed and evaluated perceptually with regard to binaural synthesis. Applied to the binaural impulse response pairs of binaural synthesis systems, AAS is shown to be able to affect sound color perception and auditory localization. The achievable degree of inaudible AAS depends on the simulated environment and decays globally with growing reverberation time. If the amount of AAS is increased beyond the perceptibility threshold, the sound color perception is affected prior to the auditory localization. Regarding the hearing sensation position, the horizontal localization appears most stable, whereas vertical position, distance, and width ratings are likely affected by AAS.

The concept of *loudness transfer functions* (LTFs) is introduced to allow for the auditory evaluation and technically motivated visualization of the loudness transmission characteristics of audio reproduction systems. An LTF represents the frequency dependent level difference at equal loudness between an audio reproduction system and the corresponding reference scene, as for example acquired in the perceptual free-field or diffuse-field equalization process of headphones. LTFs can indicate the sound color deteriorations to be expected for the systems under consideration, especially for steady state sounds. Frequency independence of the LTF with regard to the reference scene can be considered a quality criterion for virtual acoustics systems regarding sound coloration.

In order to provide a tool for the quantitative auditory quality evaluation of virtual acoustics systems, a procedure referred to as *quality assessment by just noticeable sound changes* is proposed. Reproducing the just noticeable sound changes of the reference scene is a necessary requirement for transparent audio reproduction systems. Therefore, the just noticeable sound changes provided by the transmission system under consideration quantify deviations from the reference scene and provide a measure allowing for the specification of perceptually optimized system performance requirements. Adaptive methods for just noticeable sound change measurements regarding directional and distance hearing are proposed, and the overall procedure is validated by the example of the horizontal angular resolution required for a specific dynamic binaural synthesis system, implemented with blocked auditory canal entrance recording, which is determined to about 0.5°.

Theoretical advancements derived in this thesis include a *categorization scheme for virtual acoustics procedures* distinguishing physically and psychoacoustically motivated approaches. In addition, *stimulus definitions* are proposed for the employment of virtual acoustics systems for the audio playback in hearing research and for the evaluation of virtual acoustics systems by psychoacoustic methods. Furthermore, the necessity of a nonindividual stimulus definition for conventional headphone reproduction is identified. On that basis, refined application ranges and instrumental measurement procedures for the free-field and diffuse-field equalization of headphones are proposed and verified.

The *enhancements of binaural synthesis* introduced in this work are derived from system-theoretic descriptions of a static reference scene for three recording methods: probe microphone recording with the probe tube tips in the auditory canals close to the eardrums, miniature microphone recording at the entrances to the blocked auditory canals, and artificial head recording. For each recording method, the assumptions enabling the derivation of the respective binaural synthesis theory and associated procedural shortcomings are identified. The resulting system-theoretic formulations of static binaural synthesis procedures show the possibilities and limitations of the recording methods in principle and allow for a comparison of all possible system configurations, for example with regard to the selection of the procedure suited best for a specific application or the validation of listening experiment results. All configurations possible with the discussed recording methods are included in order to identify and describe theoretically suboptimal procedures, which may occur in practical applications due to implementation constraints.

Based on the theoretical framework, a *binaural synthesis quality criterion* (BSQC) is proposed, providing an artificial head authenticity measure for binaurally synthesized ear signals, the sound pressure signals detected by the eardrums when listening to a binaural synthesis system. The BSQC can be considered a tool allowing for the artificial head verification and the instrumental comparison of binaural synthesis systems.

Extensions of the binaural synthesis theory necessary for approximately re-synthesizing temporally varying listening environments are formulated using the methods of linear dynamic systems. With regard to implementations, a procedure for determining the grid resolution required for transparent dynamic binaural synthesis is introduced, and the validity of the assumptions enabling the derivation of the binaural synthesis theory is addressed. The discussion includes an inter- and intra-individual variability analysis of the *transfer characteristics of three circum-aural headphones*. For all specimens, the inter-individual magnitude spectrum variability exceeds the average intra-individual variability due to headphone repositioning. In general, inter- and intra-individual variability values up to 10 dB in level and 0.5 ms in group delay occur at frequencies above about 6 kHz. Further, the average magnitude equalization of blocked auditory canal entrance headphone transfer functions is shown to provide equalized individual blocked auditory canal entrance headphone transfer functions on average frequency independent within ±3 dB, comparable to the deviations of the average results achieved with individual magnitude equalization.

For binaural synthesis implemented with recording at the blocked auditory canal entrance, the system-theoretic binaural synthesis framework derived here indicates a possible influence of the specific headphones employed on the resulting ear signals. In order to quantify the expected influence, a *blocked auditory canal headphone selection criterion* (HPSC) is derived, which predicts, based on four artificial head transfer function measurements, the headphone contribution to the deviations between the ear signals of a reference scene and the corresponding binaural synthesis with blocked auditory canal entrance recording. Artificial head evaluation of the HPSC indicates deviations between the HPSC prediction and the binaurally synthesized ear signal spectra below the accuracy of the measurement procedure in the range of ±0.5 dB for the magnitude spectrum and ±0.5 ms for the group delay characteristics.

With regard to applications, *perceptual consequences* of violations of the assumptions that allow for the derivation of the binaural synthesis theory are addressed for the different recording methods by loudness comparisons between binaurally synthesized and the corresponding real scenarios. The loudness transfer is shown to be affected by decaying binaural synthesis quality prior to the auditory localization and is therefore considered the more critical perceptual quality measure. The loudness comparison experiments confirm with the procedural accuracy of ±2 dB the validity of the theoretical framework introduced and illustrate its benefit for practical applications of binaural synthesis systems. In detail, individually equalized binaural synthesis is shown to be able to frequency independently elicit the reference scene loudness of narrow-band noise impulses in a non-acoustically not modified reference scene. Equal loudness for the average listener is achieved with binaural synthesis, implemented based on individual blocked auditory canal entrance recording with headphones selected according to the HPSC. Less individualized binaural synthesis systems implemented according to the proposed framework can provide correct low-frequency loudness transfer, while deviations occur above an upper limiting frequency, depending on the specific implementation, in the range of 5 kHz. Regarding the loudness transfer, negligible differences occurred in the configuration studied between the static and dynamic binaural synthesis systems employed for seated subjects facing the sound source. In summary, by the discussed loudness adjustment experiments, the virtual acoustics procedure *binaural synthesis is validated using methods of hearing research,* affirming interrelations of both fields of research.

In combination, the theoretical and experimental results of this work provide an explanation of the effect frequently referred to as *the case of the missing 6 dB*, a deviation between the auditory canal levels at equal loudness in headphone versus loudspeaker reproduction. The loudness adjustment results presented in this thesis show that the effect disappears if identical sound pressure time signals at the eardrums are targeted, for example by individual binaural synthesis, instead of equal auditory canal levels.

Concluding, a schematic working model based on Theile's association principle is proposed, summarizing the findings of this thesis by describing the buildup of hearing sensations in accordance to the presented results especially with regard to auditory localization and loudness. The model suggests dependencies between auditory localization and loudness respectively sound color. These *indications of the functionality of the human hearing system,* detected by the *virtual acoustics procedure binaural synthesis*, further confirm the benefits of utilizing *interrelations of virtual acoustics and hearing research.*

A List of Abbreviations

2-AFC	two-alternative forced choice
3-AFC	three-alternative forced choice
AAA	auditory-adapted analysis
AAS	auditory-adapted exponential transfer function smoothing
AH	artificial head
ANOVA	analysis of variance
BIRP	binaural impulse response pair
BS	binaural synthesis
BSQC	binaural synthesis quality criterion
BTFP	binaural transfer function pair
CBR	critical-band rate
CBW	critical bandwidth
CTC	crosstalk cancellation
DFT	discrete Fourier transform
ESS	exponential sine sweep
FEC	free-air equivalent coupling to the ear
FIR	finite impulse response
FTT	Fourier-t transform
HP	headphone
HPIR	headphone impulse response
HPSC	blocked auditory canal headphone selection criterion
HPTF	headphone transfer function
ILD	interaural level difference
IR	impulse response
ITD	interaural time difference
JNDEG	just noticeable degradation
JNSC	just noticeable sound change

LS	loudspeaker
LTF	loudness transfer function
LTI	linear time-invariant
MAA	minimum audible angle
MAD	minimum audible distance
MAMA	minimum audible movement angle
NBN	narrow-band noise
PDR	pressure division ratio
SNR	signal to noise ratio
TF	transfer function
UEN	uniform exciting noise
VA	virtual acoustics
WFS	wave field synthesis

B List of Symbols, Subscripts, and Superscripts

h, H	impulse response, transfer function
p, P	sound pressure, sound pressure spectrum
s, S	digital sample sequence, DFT spectrum
w, W	window function, corresponding spectral kernel
f, Δf, Δf_{G}	frequency, bandwidth, critical bandwidth
$\mathbf{x}$	position vector
z	critical-band rate
Z	impedance

abs	absolute value
ad	analog to digital
ah, ahm	artificial head, artificial head microphone
als, am	loudspeaker amplifier, microphone amplifier
avg	average
b	blocked auditory canal
bs, bsqc	binaural synthesis, binaural synthesis quality criterion
da	digital to analog
e	ear signal
eq	equalization filter
h	superscript: under the headphone subscript: head
hp, hptf	headphone, headphone transfer function
hpsc	headphone selection criterion
i	input
ind	individual
ls	loudspeaker
m	miniature microphone
mic	microphone
nind	nonindividual
o	output
play	playback situation
pm	probe microphone
rec	recording situation
ref	reference scene
ne	temporary non-equalized binaural synthesis
ver	artificial head verification of binaural synthesis

C Binaural Synthesis with Headphone Reference Scene

Binaural synthesis (BS) with headphone reference scene according to definition 30 differs from conventional BS in that the listener wears inactive headphones (HPs) in reference scene and recording situation. Consequently, the binaural impulse response pairs (BIRPs) are recorded through the HPs. This procedure is advantageous for comparisons of BS and reference scene, since both scenarios include the HPs. In this section, the system-theoretic basis is discussed based on the conventional BS theory derived in chapter 4.

C.1 Headphone Reference Scene

The propagation parts of the sound paths in the headphone reference scene are defined between the loudspeaker (LS) input and the sound pressure signals at the eardrums by

$$\mathbf{H}^{\mathrm{ind,h}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{e}}}(\mathbf{x}_{\mathrm{h}_{\mathrm{ref}}},\mathbf{x}_{\mathrm{ls}_{\mathrm{ref}}},\mathbf{x}_{\mathrm{hp}_{\mathrm{ref}}})=\frac{\mathbf{P}^{\mathrm{ind,h}}_{\mathrm{e}}(\mathbf{x}_{\mathrm{h}_{\mathrm{ref}}},\mathbf{x}_{\mathrm{ls}_{\mathrm{ref}}},\mathbf{x}_{\mathrm{hp}_{\mathrm{ref}}})}{U_{\mathrm{ls}}}. \tag{C.1}$$

Combined, equations 2.7 and C.1 result in the headphone reference scene ear signal spectra

$$\mathbf{P}^{\mathrm{ind,h}}_{\mathrm{e}}(\mathbf{x}_{\mathrm{h}_{\mathrm{ref}}},\mathbf{x}_{\mathrm{ls}_{\mathrm{ref}}},\mathbf{x}_{\mathrm{hp}_{\mathrm{ref}}})=S_{\mathrm{ls}}\cdot H_{\mathrm{o}_{\mathrm{ref}}}\cdot\mathbf{H}^{\mathrm{ind,h}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{e}}}(\mathbf{x}_{\mathrm{h}_{\mathrm{ref}}},\mathbf{x}_{\mathrm{ls}_{\mathrm{ref}}},\mathbf{x}_{\mathrm{hp}_{\mathrm{ref}}}). \tag{C.2}$$

The transfer functions (TFs) relating the digital sequence driving the LS to the ear signals in the headphone reference scene are defined by

$$\begin{aligned}\mathbf{H}^{\mathrm{ind,h}}_{\mathrm{ref}}(\mathbf{x}_{\mathrm{h}_{\mathrm{ref}}},\mathbf{x}_{\mathrm{ls}_{\mathrm{ref}}},\mathbf{x}_{\mathrm{hp}_{\mathrm{ref}}})&=\frac{\mathbf{P}^{\mathrm{ind,h}}_{\mathrm{e}}(\mathbf{x}_{\mathrm{h}_{\mathrm{ref}}},\mathbf{x}_{\mathrm{ls}_{\mathrm{ref}}},\mathbf{x}_{\mathrm{hp}_{\mathrm{ref}}})}{S_{\mathrm{ls}}}\\&=H_{\mathrm{o}_{\mathrm{ref}}}\cdot\mathbf{H}^{\mathrm{ind,h}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{e}}}(\mathbf{x}_{\mathrm{h}_{\mathrm{ref}}},\mathbf{x}_{\mathrm{ls}_{\mathrm{ref}}},\mathbf{x}_{\mathrm{hp}_{\mathrm{ref}}}).\end{aligned} \tag{C.3}$$

Probe Microphone Approximation The relation between the sound pressure spectra at probe microphones in the auditory canals and the LS driving spectrum is defined by

$$\mathbf{H}^{\mathrm{ind,h}}_{\mathrm{ref}_{\mathrm{pm}}}(\mathbf{x}_{\mathrm{h}_{\mathrm{ref}}},\mathbf{x}_{\mathrm{ls}_{\mathrm{ref}}},\mathbf{x}_{\mathrm{pm}_{\mathrm{ref}}},\mathbf{x}_{\mathrm{hp}_{\mathrm{ref}}})=\frac{\mathbf{P}^{\mathrm{ind,h}}_{\mathrm{pm}}(\mathbf{x}_{\mathrm{h}_{\mathrm{ref}}},\mathbf{x}_{\mathrm{ls}_{\mathrm{ref}}},\mathbf{x}_{\mathrm{pm}_{\mathrm{ref}}},\mathbf{x}_{\mathrm{hp}_{\mathrm{ref}}})}{S_{\mathrm{ls}}}. \tag{C.4}$$

If the sound pressure signals at the microphones are assumed to represent the ear signals, the headphone reference scene TFs can be approximated using equations 2.6 and C.4 by

$$\begin{aligned}&\mathbf{H}^{\mathrm{ind,h}}_{\mathrm{ref}}(\mathbf{x}_{\mathrm{h}_{\mathrm{ref}}},\mathbf{x}_{\mathrm{ls}_{\mathrm{ref}}},\mathbf{x}_{\mathrm{hp}_{\mathrm{ref}}})\approx\mathbf{H}^{\mathrm{ind,h}}_{\mathrm{ref}_{\mathrm{pm}}}(\mathbf{x}_{\mathrm{h}_{\mathrm{ref}}},\mathbf{x}_{\mathrm{ls}_{\mathrm{ref}}},\mathbf{x}_{\mathrm{pm}_{\mathrm{ref}}},\mathbf{x}_{\mathrm{hp}_{\mathrm{ref}}})\\&=H_{\mathrm{o}_{\mathrm{ref}}}\cdot\mathbf{H}^{\mathrm{ind,h}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{pm}}}(\mathbf{x}_{\mathrm{h}_{\mathrm{ref}}},\mathbf{x}_{\mathrm{ls}_{\mathrm{ref}}},\mathbf{x}_{\mathrm{pm}_{\mathrm{ref}}},\mathbf{x}_{\mathrm{hp}_{\mathrm{ref}}})=\frac{\mathbf{S}^{\mathrm{ind,h}}_{\mathrm{pm}}(\mathbf{x}_{\mathrm{h}_{\mathrm{ref}}},\mathbf{x}_{\mathrm{ls}_{\mathrm{ref}}},\mathbf{x}_{\mathrm{pm}_{\mathrm{ref}}},\mathbf{x}_{\mathrm{hp}_{\mathrm{ref}}})}{S_{\mathrm{ls}}\cdot\mathbf{H}_{\mathrm{pm}}\cdot\mathbf{H}_{\mathrm{i}_{\mathrm{pm}}}}.\end{aligned} \tag{C.5}$$

Artificial Head Approximation The relation of the LS driving and the sound pressure spectra at artificial head (AH) microphones in the headphone reference scene is defined by

$$\mathbf{H}^{\mathrm{ah,h}}_{\mathrm{ref_{ahm}}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{hp_{ref}}}) = \frac{\mathbf{P}^{\mathrm{ah,h}}_{\mathrm{ahm}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{hp_{ref}}})}{S_{\mathrm{ls}}}, \tag{C.6}$$

and the TFs given by equation C.3 are approximated with equations C.6 and 2.6 by

$$\begin{aligned} &\mathbf{H}^{\mathrm{ind,h}}_{\mathrm{ref}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{hp_{ref}}}) \approx \mathbf{H}^{\mathrm{ah,h}}_{\mathrm{ref_{ahm}}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{hp_{ref}}}) \\ &\quad = H_{\mathrm{o_{ref}}} \cdot \mathbf{H}^{\mathrm{ah,h}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{hp_{ref}}}) = \frac{\mathbf{S}^{\mathrm{ah,h}}_{\mathrm{ahm}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{hp_{ref}}})}{S_{\mathrm{ls}} \cdot \mathbf{H}_{\mathrm{i_{ah}}} \cdot \mathbf{H}_{\mathrm{ahm}}}. \end{aligned} \tag{C.7}$$

C.2 Recording Situation

Using the TFs of the sound paths from the LS input to the microphone output voltages

$$\mathbf{H}^{\mathrm{ind,h}}_{u_{\mathrm{ls}},\mathbf{u}_{\mathrm{mic}}}(\mathbf{x}_{\mathrm{h_{rec}}}, \mathbf{x}_{\mathrm{ls_{rec}}}, \mathbf{x}_{\mathrm{mic_{rec}}}, \mathbf{x}_{\mathrm{hp_{rec}}}) = \frac{\mathbf{U}^{\mathrm{ind,h}}_{\mathrm{mic}}(\mathbf{x}_{\mathrm{h_{rec}}}, \mathbf{x}_{\mathrm{ls_{rec}}}, \mathbf{x}_{\mathrm{mic_{rec}}}, \mathbf{x}_{\mathrm{hp_{rec}}})}{U_{\mathrm{ls}}}, \tag{C.8}$$

the recording situation TFs are given using equations 2.5, 2.7, and C.8 by

$$\begin{aligned} \mathbf{H}^{\mathrm{ind,h}}_{\mathrm{rec_{mic}}}(\mathbf{x}_{\mathrm{h_{rec}}}, \mathbf{x}_{\mathrm{ls_{rec}}}, \mathbf{x}_{\mathrm{mic_{rec}}}, \mathbf{x}_{\mathrm{hp_{rec}}}) &= \frac{\mathbf{S}^{\mathrm{ind,h}}_{\mathrm{mic}}(\mathbf{x}_{\mathrm{h_{rec}}}, \mathbf{x}_{\mathrm{ls_{rec}}}, \mathbf{x}_{\mathrm{mic_{rec}}}, \mathbf{x}_{\mathrm{hp_{rec}}})}{S_{\mathrm{ls}}} \\ &= H_{\mathrm{o_{rec}}} \cdot \mathbf{H}^{\mathrm{ind,h}}_{u_{\mathrm{ls}},\mathbf{u}_{\mathrm{mic}}}(\mathbf{x}_{\mathrm{h_{rec}}}, \mathbf{x}_{\mathrm{ls_{rec}}}, \mathbf{x}_{\mathrm{mic_{rec}}}, \mathbf{x}_{\mathrm{hp_{rec}}}) \cdot \mathbf{H}_{\mathrm{i_{rec}}}. \end{aligned} \tag{C.9}$$

Probe Microphone Recording The TFs describing the recording situation with probe microphones are given using equations 2.6 and C.9 by

$$\begin{aligned} &\mathbf{H}^{\mathrm{ind,h}}_{\mathrm{rec_{pm}}}(\mathbf{x}_{\mathrm{h_{rec}}}, \mathbf{x}_{\mathrm{ls_{rec}}}, \mathbf{x}_{\mathrm{pm_{rec}}}, \mathbf{x}_{\mathrm{hp_{rec}}}) = \\ &\qquad = H_{\mathrm{o_{rec}}} \cdot \mathbf{H}^{\mathrm{ind,h}}_{u_{\mathrm{ls}},\mathbf{u}_{\mathrm{pm}}}(\mathbf{x}_{\mathrm{h_{rec}}}, \mathbf{x}_{\mathrm{ls_{rec}}}, \mathbf{x}_{\mathrm{pm_{rec}}}, \mathbf{x}_{\mathrm{hp_{rec}}}) \cdot \mathbf{H}_{\mathrm{i_{pm}}} \\ &\qquad = H_{\mathrm{o_{rec}}} \cdot \mathbf{H}^{\mathrm{ind,h}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{pm}}}(\mathbf{x}_{\mathrm{h_{rec}}}, \mathbf{x}_{\mathrm{ls_{rec}}}, \mathbf{x}_{\mathrm{pm_{rec}}}, \mathbf{x}_{\mathrm{hp_{rec}}}) \cdot \mathbf{H}_{\mathrm{pm}} \cdot \mathbf{H}_{\mathrm{i_{pm}}}, \end{aligned} \tag{C.10}$$

approximately identical to the reference scene TFs (equation C.3), if identical positions are assumed (cf. equations C.5 and C.9). This is formulated by

$$\mathbf{H}^{\mathrm{ind,h}}_{\mathrm{rec_{pm}}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{pm_{rec}}}, \mathbf{x}_{\mathrm{hp_{rec}}}) \approx \mathbf{H}^{\mathrm{ind,h}}_{\mathrm{ref}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{hp_{ref}}}) \cdot \mathbf{H}_{\mathrm{pm}} \cdot \mathbf{H}_{\mathrm{i_{pm}}}. \tag{C.11}$$

Blocked Auditory Canal Recording The TFs describing the recording with miniature microphones in the blocked auditory canals are with equations 2.6 and C.9 given by

$$\begin{aligned} &\mathbf{H}^{\mathrm{ind,b,h}}_{\mathrm{rec_{m}}}(\mathbf{x}_{\mathrm{h_{rec}}}, \mathbf{x}_{\mathrm{ls_{rec}}}, \mathbf{x}_{\mathrm{m_{rec}}}, \mathbf{x}_{\mathrm{hp_{rec}}}) = \\ &\qquad = H_{\mathrm{o_{rec}}} \cdot \mathbf{H}^{\mathrm{ind,b,h}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{m}}}(\mathbf{x}_{\mathrm{h_{rec}}}, \mathbf{x}_{\mathrm{ls_{rec}}}, \mathbf{x}_{\mathrm{m_{rec}}}, \mathbf{x}_{\mathrm{hp_{rec}}}) \cdot \mathbf{H}_{\mathrm{m}} \cdot \mathbf{H}_{\mathrm{i_{m}}}. \end{aligned} \tag{C.12}$$

Comparing equations C.3 and C.12 and assuming identical HP positions results in

$$\begin{aligned}&\mathbf{H}_{\mathrm{rec_m}}^{\mathrm{ind,b,h}}(\mathbf{x}_{\mathrm{h_{ref}}},\mathbf{x}_{\mathrm{ls_{ref}}},\mathbf{x}_{\mathrm{m_{rec}}},\mathbf{x}_{\mathrm{hp_{rec}}}) =\\ &\quad = \mathbf{H}_{\mathrm{ref}}^{\mathrm{ind,h}}(\mathbf{x}_{\mathrm{h_{ref}}},\mathbf{x}_{\mathrm{ls_{ref}}},\mathbf{x}_{\mathrm{hp_{ref}}}) \cdot \frac{\mathbf{H}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{m}}}^{\mathrm{ind,b,h}}(\mathbf{x}_{\mathrm{h_{ref}}},\mathbf{x}_{\mathrm{ls_{ref}}},\mathbf{x}_{\mathrm{m_{rec}}},\mathbf{x}_{\mathrm{hp_{rec}}})}{\mathbf{H}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{e}}}^{\mathrm{ind,h}}(\mathbf{x}_{\mathrm{h_{ref}}},\mathbf{x}_{\mathrm{ls_{ref}}},\mathbf{x}_{\mathrm{hp_{ref}}})} \cdot \mathbf{H}_{\mathrm{m}} \cdot \mathbf{H}_{\mathrm{i_m}}.\end{aligned} \tag{C.13}$$

Artificial Head Recording With equations 2.6 and C.9, AH recording is described by

$$\mathbf{H}_{\mathrm{rec_{ahm}}}^{\mathrm{ah,h}}(\mathbf{x}_{\mathrm{h_{rec}}},\mathbf{x}_{\mathrm{ls_{rec}}},\mathbf{x}_{\mathrm{hp_{rec}}}) = H_{\mathrm{o_{rec}}} \cdot \mathbf{H}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{ahm}}}^{\mathrm{ah,h}}(\mathbf{x}_{\mathrm{h_{rec}}},\mathbf{x}_{\mathrm{ls_{rec}}},\mathbf{x}_{\mathrm{hp_{rec}}}) \cdot \mathbf{H}_{\mathrm{ahm}} \cdot \mathbf{H}_{\mathrm{i_{ah}}}. \tag{C.14}$$

By comparison of equations C.7 and C.14 under the assumption of identical HP positions in recording situation and reference scene,

$$\mathbf{H}_{\mathrm{rec_{ahm}}}^{\mathrm{ah,h}}(\mathbf{x}_{\mathrm{h_{ref}}},\mathbf{x}_{\mathrm{ls_{ref}}},\mathbf{x}_{\mathrm{hp_{ref}}}) = \mathbf{H}_{\mathrm{ref_{ahm}}}^{\mathrm{ah,h}}(\mathbf{x}_{\mathrm{h_{ref}}},\mathbf{x}_{\mathrm{ls_{ref}}},\mathbf{x}_{\mathrm{hp_{ref}}}) \cdot \mathbf{H}_{\mathrm{ahm}} \cdot \mathbf{H}_{\mathrm{i_{ah}}} \tag{C.15}$$

is derived, formulated in relation to the reference scene using equations C.3 and C.14 by

$$\begin{aligned}&\mathbf{H}_{\mathrm{rec_{ahm}}}^{\mathrm{ah,h}}(\mathbf{x}_{\mathrm{h_{ref}}},\mathbf{x}_{\mathrm{ls_{ref}}},\mathbf{x}_{\mathrm{hp_{ref}}}) =\\ &\quad = \mathbf{H}_{\mathrm{ref}}^{\mathrm{ind,h}}(\mathbf{x}_{\mathrm{h_{ref}}},\mathbf{x}_{\mathrm{ls_{ref}}},\mathbf{x}_{\mathrm{hp_{ref}}}) \cdot \frac{\mathbf{H}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{ahm}}}^{\mathrm{ah,h}}(\mathbf{x}_{\mathrm{h_{ref}}},\mathbf{x}_{\mathrm{ls_{ref}}},\mathbf{x}_{\mathrm{hp_{ref}}})}{\mathbf{H}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{e}}}^{\mathrm{ind,h}}(\mathbf{x}_{\mathrm{h_{ref}}},\mathbf{x}_{\mathrm{ls_{ref}}},\mathbf{x}_{\mathrm{hp_{ref}}})} \cdot \mathbf{H}_{\mathrm{ahm}} \cdot \mathbf{H}_{\mathrm{i_{ah}}}.\end{aligned} \tag{C.16}$$

C.3 Playback Situation and Headphone Transfer Functions

The BS playback situation is independent of the reference scene. For that reason, section 4.4 holds true also for BS with headphone reference scene. In particular, the headphone transfer functions (HPTFs) are valid also for the HP reference scene.

C.4 Non-Equalized Binaural Synthesis

Using equations 4.21 and C.9 and assuming $\mathbf{x}_{\mathrm{h_{rec}}} = \mathbf{x}_{\mathrm{h_{ref}}}$ and $\mathbf{x}_{\mathrm{ls_{rec}}} = \mathbf{x}_{\mathrm{ls_{ref}}}$, the ear signals of the non-equalized BS situation with headphone reference scene are computed to

$$\begin{aligned}&\mathbf{P}_{\mathrm{e_{ne}}}^{\mathrm{ind,h}}(\mathbf{x}_{\mathrm{h_{ref}}},\mathbf{x}_{\mathrm{ls_{ref}}},\mathbf{x}_{\mathrm{mic_{rec}}},\mathbf{x}_{\mathrm{hp_{rec}}},\mathbf{x}_{\mathrm{hp_{play}}}) =\\ &\qquad = S_{\mathrm{ls}} \cdot \mathbf{H}_{\mathrm{rec_{mic}}}^{\mathrm{ind,h}}(\mathbf{x}_{\mathrm{h_{ref}}},\mathbf{x}_{\mathrm{ls_{ref}}},\mathbf{x}_{\mathrm{mic_{rec}}},\mathbf{x}_{\mathrm{hp_{rec}}}) \cdot \mathbf{H}_{\mathrm{play}}^{\mathrm{ind,h}}(\mathbf{x}_{\mathrm{hp_{play}}}).\end{aligned} \tag{C.17}$$

In the following, different BS situations are described by the TFs

$$\begin{aligned}&\mathbf{H}_{\mathrm{ne}}^{\mathrm{ind,h}}(\mathbf{x}_{\mathrm{h_{ref}}},\mathbf{x}_{\mathrm{ls_{ref}}},\mathbf{x}_{\mathrm{mic_{rec}}},\mathbf{x}_{\mathrm{hp_{rec}}},\mathbf{x}_{\mathrm{hp_{play}}}) =\\ &\qquad = \frac{\mathbf{P}_{\mathrm{e_{ne}}}^{\mathrm{ind,h}}(\mathbf{x}_{\mathrm{h_{ref}}},\mathbf{x}_{\mathrm{ls_{ref}}},\mathbf{x}_{\mathrm{mic_{rec}}},\mathbf{x}_{\mathrm{hp_{rec}}},\mathbf{x}_{\mathrm{hp_{play}}})}{S_{\mathrm{ls}}}\\ &\qquad = \mathbf{H}_{\mathrm{rec_{mic}}}^{\mathrm{ind,h}}(\mathbf{x}_{\mathrm{h_{ref}}},\mathbf{x}_{\mathrm{ls_{ref}}},\mathbf{x}_{\mathrm{mic_{rec}}},\mathbf{x}_{\mathrm{hp_{rec}}}) \cdot \mathbf{H}_{\mathrm{play}}^{\mathrm{ind,h}}(\mathbf{x}_{\mathrm{hp_{play}}}).\end{aligned} \tag{C.18}$$

Non-Equalized Human Head Playback Combining equations 4.21 and C.11 results in

$$\begin{aligned}&\mathbf{H}^{\mathrm{ind,h}}_{\mathrm{ne_{pm}}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{pm_{rec}}}, \mathbf{x}_{\mathrm{hp_{rec}}}, \mathbf{x}_{\mathrm{hp_{play}}}) = H_{\mathrm{o_{ref}}}\cdot\\ &\quad\cdot \mathbf{H}^{\mathrm{ind,h}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{pm}}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{pm_{rec}}}, \mathbf{x}_{\mathrm{hp_{rec}}}) \cdot \mathbf{H}_{\mathrm{pm}} \cdot \mathbf{H}_{\mathrm{i_{pm}}} \cdot \mathbf{H}_{\mathrm{o_{hp}}} \cdot \mathbf{H}^{\mathrm{ind,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{e}}}(\mathbf{x}_{\mathrm{hp_{play}}}),\end{aligned} \tag{C.19}$$

connecting the source signal and the ear signal spectra for probe microphone recording and human head playback. Assuming the sound pressures at the *probe microphone* positions during the recording to represent the ear signals allows for the approximation

$$\begin{aligned}&\mathbf{H}^{\mathrm{ind,h}}_{\mathrm{ne_{pm}}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{pm_{rec}}}, \mathbf{x}_{\mathrm{hp_{rec}}}, \mathbf{x}_{\mathrm{hp_{play}}}) \approx\\ &\quad\approx \mathbf{H}^{\mathrm{ind,h}}_{\mathrm{ref}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{hp_{ref}}}) \cdot \mathbf{H}_{\mathrm{pm}} \cdot \mathbf{H}_{\mathrm{i_{pm}}} \cdot \mathbf{H}_{\mathrm{o_{hp}}} \cdot \mathbf{H}^{\mathrm{ind,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{e}}}(\mathbf{x}_{\mathrm{hp_{play}}}).\end{aligned} \tag{C.20}$$

Blocked auditory canal recording and human playback results (equations 4.21 and C.13) in

$$\begin{aligned}&\mathbf{H}^{\mathrm{ind,h}}_{\mathrm{ne_{m}}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{m_{rec}}}, \mathbf{x}_{\mathrm{hp_{rec}}}, \mathbf{x}_{\mathrm{hp_{play}}}) = \mathbf{H}^{\mathrm{ind,h}}_{\mathrm{ref}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{hp_{ref}}})\cdot\\ &\quad\cdot \frac{\mathbf{H}^{\mathrm{ind,b,h}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{m}}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{m_{rec}}}, \mathbf{x}_{\mathrm{hp_{rec}}})}{\mathbf{H}^{\mathrm{ind,h}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{e}}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{hp_{ref}}})} \cdot \mathbf{H}_{\mathrm{m}} \cdot \mathbf{H}_{\mathrm{i_{m}}} \cdot \mathbf{H}_{\mathrm{o_{hp}}} \cdot \mathbf{H}^{\mathrm{ind,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{e}}}(\mathbf{x}_{\mathrm{hp_{play}}}).\end{aligned} \tag{C.21}$$

Artificial head recording and human head playback is given (equations 4.21 and C.16) by

$$\begin{aligned}&\mathbf{H}^{\mathrm{ind,h}}_{\mathrm{ne_{ah}}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{hp_{rec}}}, \mathbf{x}_{\mathrm{hp_{play}}}) = \mathbf{H}^{\mathrm{ind,h}}_{\mathrm{ref}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{hp_{ref}}})\cdot\\ &\quad\cdot \frac{\mathbf{H}^{\mathrm{ah,h}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{hp_{rec}}})}{\mathbf{H}^{\mathrm{ind,h}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{e}}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{hp_{ref}}})} \cdot \mathbf{H}_{\mathrm{ahm}} \cdot \mathbf{H}_{\mathrm{i_{ah}}} \cdot \mathbf{H}_{\mathrm{o_{hp}}} \cdot \mathbf{H}^{\mathrm{ind,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{e}}}(\mathbf{x}_{\mathrm{hp_{play}}}).\end{aligned} \tag{C.22}$$

Non-Equalized Artificial Head Playback If recording using *probe microphones*, a combination of equations 2.7, 4.21, and C.11 describes the AH playback situation by

$$\begin{aligned}&\mathbf{H}^{\mathrm{ah,h}}_{\mathrm{ne_{pm}}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{pm_{rec}}}, \mathbf{x}_{\mathrm{hp_{rec}}}, \mathbf{x}_{\mathrm{hp_{play}}}) = H_{\mathrm{o_{ref}}}\cdot\\ &\quad\cdot \mathbf{H}^{\mathrm{ind,h}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{pm}}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{pm_{rec}}}, \mathbf{x}_{\mathrm{hp_{rec}}}) \cdot \mathbf{H}_{\mathrm{pm}} \cdot \mathbf{H}_{\mathrm{i_{pm}}} \cdot \mathbf{H}_{\mathrm{o_{hp}}} \cdot \mathbf{H}^{\mathrm{ah,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{hp_{play}}}),\end{aligned} \tag{C.23}$$

given in relation to the AH headphone reference scene based on equation C.7 by

$$\begin{aligned}&\mathbf{H}^{\mathrm{ah,h}}_{\mathrm{ne_{pm}}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{pm_{rec}}}, \mathbf{x}_{\mathrm{hp_{rec}}}, \mathbf{x}_{\mathrm{hp_{play}}}) = \mathbf{H}^{\mathrm{ah,h}}_{\mathrm{ref_{ahm}}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{hp_{ref}}})\cdot\\ &\quad\cdot \frac{\mathbf{H}^{\mathrm{ind,h}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{pm}}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{pm_{rec}}}, \mathbf{x}_{\mathrm{hp_{rec}}})}{\mathbf{H}^{\mathrm{ah,h}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{hp_{ref}}})} \cdot \mathbf{H}_{\mathrm{pm}} \cdot \mathbf{H}_{\mathrm{i_{pm}}} \cdot \mathbf{H}_{\mathrm{o_{hp}}} \cdot \mathbf{H}^{\mathrm{ah,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{hp_{play}}}).\end{aligned} \tag{C.24}$$

The AH playback of *blocked auditory canal* recordings given (equations 2.7, 4.21, C.12) by

$$\begin{aligned}&\mathbf{H}^{\mathrm{ah,h}}_{\mathrm{ne_{m}}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{m_{rec}}}, \mathbf{x}_{\mathrm{hp_{rec}}}, \mathbf{x}_{\mathrm{hp_{play}}}) = H_{\mathrm{o_{ref}}}\cdot\\ &\quad\cdot \mathbf{H}^{\mathrm{ind,b,h}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{m}}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{m_{rec}}}, \mathbf{x}_{\mathrm{hp_{rec}}}) \cdot \mathbf{H}_{\mathrm{m}} \cdot \mathbf{H}_{\mathrm{i_{m}}} \cdot \mathbf{H}_{\mathrm{o_{hp}}} \cdot \mathbf{H}^{\mathrm{ah,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{hp_{play}}})\end{aligned} \tag{C.25}$$

is related to the headphone reference scene using equation C.7, resulting in

$$\begin{aligned}\mathbf{H}^{\mathrm{ah,h}}_{\mathrm{ne_m}}(\mathbf{x}_{\mathrm{h_{ref}}},\mathbf{x}_{\mathrm{ls_{ref}}},\mathbf{x}_{\mathrm{m_{rec}}},\mathbf{x}_{\mathrm{hp_{rec}}},\mathbf{x}_{\mathrm{hp_{play}}}) &= \mathbf{H}^{\mathrm{ah,h}}_{\mathrm{ref_{ahm}}}(\mathbf{x}_{\mathrm{h_{ref}}},\mathbf{x}_{\mathrm{ls_{ref}}},\mathbf{x}_{\mathrm{hp_{ref}}})\cdot \\ \cdot\,\frac{\mathbf{H}^{\mathrm{ind,b,h}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{m}}}(\mathbf{x}_{\mathrm{h_{ref}}},\mathbf{x}_{\mathrm{ls_{ref}}},\mathbf{x}_{\mathrm{m_{rec}}},\mathbf{x}_{\mathrm{hp_{rec}}})}{\mathbf{H}^{\mathrm{ah,h}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{h_{ref}}},\mathbf{x}_{\mathrm{ls_{ref}}},\mathbf{x}_{\mathrm{hp_{ref}}})} &\cdot \mathbf{H}_{\mathrm{m}}\cdot\mathbf{H}_{\mathrm{i_m}}\cdot\mathbf{H}_{\mathrm{o_{hp}}}\cdot\mathbf{H}^{\mathrm{ah,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{hp_{play}}}).\end{aligned}\tag{C.26}$$

With equations 4.21 and C.15, the AH playback of *artificial head* recordings is given by

$$\begin{aligned}&\mathbf{H}^{\mathrm{ah,h}}_{\mathrm{ne_{ah}}}(\mathbf{x}_{\mathrm{h_{ref}}},\mathbf{x}_{\mathrm{ls_{ref}}},\mathbf{x}_{\mathrm{hp_{rec}}},\mathbf{x}_{\mathrm{hp_{play}}}) = \\ &\quad= \mathbf{H}^{\mathrm{ah,h}}_{\mathrm{ahm}}(\mathbf{x}_{\mathrm{h_{ref}}},\mathbf{x}_{\mathrm{ls_{ref}}},\mathbf{x}_{\mathrm{hp_{rec}}})\cdot\mathbf{H}_{\mathrm{ahm}}\cdot\mathbf{H}_{\mathrm{i_{ah}}}\cdot\mathbf{H}_{\mathrm{o_{hp}}}\cdot\mathbf{H}^{\mathrm{ah,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{hp_{play}}}).\end{aligned}\tag{C.27}$$

C.5 Equalization Requirements

Following assumption 2, recreating the headphone reference scene ear signals is assumed as design target for BS with headphone reference scene, formulated mathematically by

$$\mathbf{P}^{\mathrm{ind,h}}_{\mathrm{e_{bs}}}(\mathbf{x}_{\mathrm{h_{ref}}},\mathbf{x}_{\mathrm{ls_{ref}}},\mathbf{x}_{\mathrm{mic_{rec}}},\mathbf{x}_{\mathrm{hp_{rec}}},\mathbf{x}_{\mathrm{hp_{play}}}) \overset{!}{=} \mathbf{P}^{\mathrm{ind,h}}_{\mathrm{e}}(\mathbf{x}_{\mathrm{h_{ref}}},\mathbf{x}_{\mathrm{ls_{ref}}},\mathbf{x}_{\mathrm{hp_{ref}}}).\tag{C.28}$$

In order to reach this goal, equalization filters $\mathbf{h}^{\mathrm{ind,h}}_{\mathrm{eq}}$ have to be applied, which are defined combining the results of section C.4 with equations C.2 and C.17 by

$$\begin{aligned}\mathbf{H}^{\mathrm{ind,h}}_{\mathrm{ne}}(\mathbf{x}_{\mathrm{h_{ref}}},\mathbf{x}_{\mathrm{ls_{ref}}},\mathbf{x}_{\mathrm{mic_{rec}}},\mathbf{x}_{\mathrm{hp_{rec}}},\mathbf{x}_{\mathrm{hp_{play}}})\cdot\mathbf{H}^{\mathrm{ind,h}}_{\mathrm{eq}}(\mathbf{x}_{\mathrm{hp_{play}}},\mathbf{x}_{\mathrm{mic_{rec}}}) &= \\ = \mathbf{H}^{\mathrm{ind,h}}_{\mathrm{ref}}(\mathbf{x}_{\mathrm{h_{ref}}},\mathbf{x}_{\mathrm{ls_{ref}}},\mathbf{x}_{\mathrm{hp_{ref}}}).&\end{aligned}\tag{C.29}$$

Assuming identical HP positions in reference scene and recording situation ($\mathbf{x}_{\mathrm{hp_{ref}}} = \mathbf{x}_{\mathrm{hp_{rec}}}$) and the invertibility of $\mathbf{H}^{\mathrm{ind,h}}_{\mathrm{ne}}$, the equalization filters are given in general by

$$\mathbf{H}^{\mathrm{ind,h}}_{\mathrm{eq}}(\mathbf{x}_{\mathrm{hp_{play}}},\mathbf{x}_{\mathrm{mic_{rec}}}) = \frac{\mathbf{H}^{\mathrm{ind,h}}_{\mathrm{ref}}(\mathbf{x}_{\mathrm{h_{ref}}},\mathbf{x}_{\mathrm{ls_{ref}}},\mathbf{x}_{\mathrm{hp_{ref}}})}{\mathbf{H}^{\mathrm{ind,h}}_{\mathrm{ne}}(\mathbf{x}_{\mathrm{h_{ref}}},\mathbf{x}_{\mathrm{ls_{ref}}},\mathbf{x}_{\mathrm{mic_{rec}}},\mathbf{x}_{\mathrm{hp_{rec}}},\mathbf{x}_{\mathrm{hp_{play}}})}.\tag{C.30}$$

Equalization: Human Head Playback, Probe Microphone Recording The equalization requirements for BS with headphone reference scene, probe microphone recording, and human head playback are given using equations C.20 and C.30 by

$$\mathbf{H}^{\mathrm{ind,h}}_{\mathrm{eq_{pm}}}(\mathbf{x}_{\mathrm{hp_{play}}},\mathbf{x}_{\mathrm{pm_{rec}}}) \approx \frac{1}{\mathbf{H}_{\mathrm{pm}}\cdot\mathbf{H}_{\mathrm{i_{pm}}}\cdot\mathbf{H}_{\mathrm{o_{hp}}}\cdot\mathbf{H}^{\mathrm{ind,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{e}}}(\mathbf{x}_{\mathrm{hp_{play}}})}.\tag{C.31}$$

With equations 4.24 and C.31, the requirements for *probe microphone recording* and human head playback are related to *probe microphone HPTFs* under assumption 4 by

$$\mathbf{H}^{\mathrm{ind,h}}_{\mathrm{eq_{pm}}}(\mathbf{x}_{\mathrm{hp_{play}}},\mathbf{x}_{\mathrm{pm_{rec}}}) \approx \frac{1}{\mathbf{H}^{\mathrm{ind,h}}_{\mathrm{hptf_{pm}}}(\mathbf{x}_{\mathrm{hp_{hptf}}},\mathbf{x}_{\mathrm{pm_{hptf}}})}.\tag{C.32}$$

Using equations 4.27 and C.31, the equalization requirements for *probe microphone recordings* and human head playback based on *blocked auditory canal HPTFs* are given by

$$\begin{aligned}\mathbf{H}^{\mathrm{ind,h}}_{\mathrm{eq_{pm}}}(\mathbf{x}_{\mathrm{hp_{play}}},\mathbf{x}_{\mathrm{pm_{rec}}}) \approx{}& \frac{1}{\mathbf{H}^{\mathrm{ind,h,b}}_{\mathrm{hptf_{m}}}(\mathbf{x}_{\mathrm{hp_{hptf}}},\mathbf{x}_{\mathrm{m_{hptf}}})}\cdot\frac{\mathbf{H}_{\mathrm{m}}\cdot\mathbf{H}_{\mathrm{i_{m}}}}{\mathbf{H}_{\mathrm{pm}}\cdot\mathbf{H}_{\mathrm{i_{pm}}}}\cdot\\ &\cdot\frac{1}{\mathbf{H}^{\mathrm{ind,h}}_{\mathbf{p}_{\mathrm{m_b}},\mathbf{p}_{\mathrm{e}},\mathrm{hp}}(\mathbf{x}_{\mathrm{hp_{play}}},\mathbf{x}_{\mathrm{hp_{hptf}}},\mathbf{x}_{\mathrm{m_{hptf}}})}.\end{aligned}\tag{C.33}$$

The combination of *probe microphone recording* and human head playback with equalization by *AH HPTFs* is described based on equations 4.29 and C.31 by

$$\mathbf{H}^{\mathrm{ind,h}}_{\mathrm{eq_{pm}}}(\mathbf{x}_{\mathrm{hp_{play}}},\mathbf{x}_{\mathrm{pm_{rec}}}) \approx \frac{1}{\mathbf{H}^{\mathrm{ah,h}}_{\mathrm{hptf_{ahm}}}(\mathbf{x}_{\mathrm{hp_{hptf}}})}\cdot\frac{\mathbf{H}^{\mathrm{ah,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{hp_{hptf}}})}{\mathbf{H}^{\mathrm{ind,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{e}}}(\mathbf{x}_{\mathrm{hp_{play}}})}\cdot\frac{\mathbf{H}_{\mathrm{ahm}}\cdot\mathbf{H}_{\mathrm{i_{ah}}}}{\mathbf{H}_{\mathrm{pm}}\cdot\mathbf{H}_{\mathrm{i_{pm}}}}.\tag{C.34}$$

Equalization: Human Head Playback, Blocked Auditory Canal Recording The human head playback of blocked auditory canal recordings is described with equations C.21 and C.30 by the TFs

$$\begin{aligned}&\mathbf{H}^{\mathrm{ind,b,h}}_{\mathrm{eq_{m}}}(\mathbf{x}_{\mathrm{hp_{play}}},\mathbf{x}_{\mathrm{m_{rec}}}) =\\ &= \frac{\mathbf{H}^{\mathrm{ind,h}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{e}}}(\mathbf{x}_{\mathrm{h_{ref}}},\mathbf{x}_{\mathrm{ls_{ref}}},\mathbf{x}_{\mathrm{hp_{ref}}})}{\mathbf{H}^{\mathrm{ind,b,h}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{m}}}(\mathbf{x}_{\mathrm{h_{ref}}},\mathbf{x}_{\mathrm{ls_{ref}}},\mathbf{x}_{\mathrm{m_{rec}}},\mathbf{x}_{\mathrm{hp_{rec}}})}\cdot\frac{1}{\mathbf{H}_{\mathrm{m}}\cdot\mathbf{H}_{\mathrm{i_{m}}}\cdot\mathbf{H}_{\mathrm{o_{hp}}}\cdot\mathbf{H}^{\mathrm{ind,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{e}}}(\mathbf{x}_{\mathrm{hp_{play}}})}.\end{aligned}\tag{C.35}$$

Human head playback and *blocked auditory canal recording* related to *probe microphone HPTFs* is given with equations 4.24 and C.35 by

$$\begin{aligned}\mathbf{H}^{\mathrm{ind,b,h}}_{\mathrm{eq_{m}}}(\mathbf{x}_{\mathrm{hp_{play}}},\mathbf{x}_{\mathrm{m_{rec}}}) ={}& \frac{1}{\mathbf{H}^{\mathrm{ind,h}}_{\mathrm{hptf_{pm}}}(\mathbf{x}_{\mathrm{hp_{hptf}}},\mathbf{x}_{\mathrm{pm_{hptf}}})}\cdot\frac{\mathbf{H}^{\mathrm{ind,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{pm}}}(\mathbf{x}_{\mathrm{hp_{hptf}}},\mathbf{x}_{\mathrm{pm_{hptf}}})}{\mathbf{H}^{\mathrm{ind,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{e}}}(\mathbf{x}_{\mathrm{hp_{play}}})}\cdot\\ &\cdot\frac{\mathbf{H}^{\mathrm{ind,h}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{e}}}(\mathbf{x}_{\mathrm{h_{ref}}},\mathbf{x}_{\mathrm{ls_{ref}}},\mathbf{x}_{\mathrm{hp_{ref}}})}{\mathbf{H}^{\mathrm{ind,b,h}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{m}}}(\mathbf{x}_{\mathrm{h_{ref}}},\mathbf{x}_{\mathrm{ls_{ref}}},\mathbf{x}_{\mathrm{m_{rec}}},\mathbf{x}_{\mathrm{hp_{rec}}})}\cdot\frac{\mathbf{H}_{\mathrm{pm}}\cdot\mathbf{H}_{\mathrm{i_{pm}}}}{\mathbf{H}_{\mathrm{m}}\cdot\mathbf{H}_{\mathrm{i_{m}}}}.\end{aligned}\tag{C.36}$$

If assumptions 4 and 6 are taken for granted, equation C.36 can be simplified to

$$\begin{aligned}&\mathbf{H}^{\mathrm{ind,b,h}}_{\mathrm{eq_{m}}}(\mathbf{x}_{\mathrm{hp_{play}}},\mathbf{x}_{\mathrm{m_{rec}}}) \approx \cdot\\ &\approx \frac{1}{\mathbf{H}^{\mathrm{ind,h}}_{\mathrm{hptf_{pm}}}(\mathbf{x}_{\mathrm{hp_{hptf}}},\mathbf{x}_{\mathrm{pm_{hptf}}})}\cdot\frac{\mathbf{H}^{\mathrm{ind,h}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{e}}}(\mathbf{x}_{\mathrm{h_{ref}}},\mathbf{x}_{\mathrm{ls_{ref}}},\mathbf{x}_{\mathrm{hp_{ref}}})}{\mathbf{H}^{\mathrm{ind,b,h}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{m}}}(\mathbf{x}_{\mathrm{h_{ref}}},\mathbf{x}_{\mathrm{ls_{ref}}},\mathbf{x}_{\mathrm{m_{rec}}},\mathbf{x}_{\mathrm{hp_{rec}}})}\cdot\frac{\mathbf{H}_{\mathrm{pm}}\cdot\mathbf{H}_{\mathrm{i_{pm}}}}{\mathbf{H}_{\mathrm{m}}\cdot\mathbf{H}_{\mathrm{i_{m}}}}.\end{aligned}\tag{C.37}$$

With equations 4.27 and C.35, the relation of the equalization for the human head playback of *blocked auditory canal recordings* to *blocked auditory canal HPTFs* is given by

$$\begin{aligned}\mathbf{H}^{\mathrm{ind,b,h}}_{\mathrm{eq_{m}}}(\mathbf{x}_{\mathrm{hp_{play}}},\mathbf{x}_{\mathrm{m_{rec}}}) ={}& \frac{1}{\mathbf{H}^{\mathrm{ind,h,b}}_{\mathrm{hptf_{m}}}(\mathbf{x}_{\mathrm{hp_{hptf}}},\mathbf{x}_{\mathrm{m_{hptf}}})}\cdot\\ \cdot\frac{\mathbf{H}^{\mathrm{ind,h}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{e}}}(\mathbf{x}_{\mathrm{h_{ref}}},\mathbf{x}_{\mathrm{ls_{ref}}},\mathbf{x}_{\mathrm{hp_{ref}}})}{\mathbf{H}^{\mathrm{ind,b,h}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{m}}}(\mathbf{x}_{\mathrm{h_{ref}}},\mathbf{x}_{\mathrm{ls_{ref}}},\mathbf{x}_{\mathrm{m_{rec}}},\mathbf{x}_{\mathrm{hp_{rec}}})}&\cdot\frac{1}{\mathbf{H}^{\mathrm{ind,h}}_{\mathbf{p}_{\mathrm{m_b}},\mathbf{p}_{\mathrm{e}},\mathrm{hp}}(\mathbf{x}_{\mathrm{hp_{play}}},\mathbf{x}_{\mathrm{hp_{hptf}}},\mathbf{x}_{\mathrm{m_{hptf}}})}.\end{aligned}\tag{C.38}$$

Defining, based on the discussion in section 4.6.1, the TFs

$$\mathbf{H}^{\mathrm{ind,h}}_{\mathbf{p}_{\mathrm{m_b}},\mathbf{p}_{\mathrm{e}},\mathrm{ls}}(\mathbf{x}_{\mathrm{m_{rec}}},\mathbf{x}_{\mathrm{hp_{rec}}},\mathbf{x}_{\mathrm{hp_{ref}}}) = \frac{\mathbf{H}^{\mathrm{ind,h}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{e}}}(\mathbf{x}_{\mathrm{h_{ref}}},\mathbf{x}_{\mathrm{ls_{ref}}},\mathbf{x}_{\mathrm{hp_{ref}}})}{\mathbf{H}^{\mathrm{ind,b,h}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{m}}}(\mathbf{x}_{\mathrm{h_{ref}}},\mathbf{x}_{\mathrm{ls_{ref}}},\mathbf{x}_{\mathrm{m_{rec}}},\mathbf{x}_{\mathrm{hp_{rec}}})}, \tag{C.39}$$

equation C.38 can be rewritten to

$$\begin{aligned}&\mathbf{H}^{\mathrm{ind,b,h}}_{\mathrm{eq_m}}(\mathbf{x}_{\mathrm{hp_{play}}},\mathbf{x}_{\mathrm{m_{rec}}}) = \\ &= \frac{1}{\mathbf{H}^{\mathrm{ind,h,b}}_{\mathrm{hptf_m}}(\mathbf{x}_{\mathrm{hp_{hptf}}},\mathbf{x}_{\mathrm{m_{hptf}}})} \cdot \frac{\mathbf{H}^{\mathrm{ind,h}}_{\mathbf{p}_{\mathrm{m_b}},\mathbf{p}_{\mathrm{e}},\mathrm{ls}}(\mathbf{x}_{\mathrm{m_{rec}}},\mathbf{x}_{\mathrm{hp_{rec}}},\mathbf{x}_{\mathrm{hp_{ref}}})}{\mathbf{H}^{\mathrm{ind,h}}_{\mathbf{p}_{\mathrm{m_b}},\mathbf{p}_{\mathrm{e}},\mathrm{hp}}(\mathbf{x}_{\mathrm{hp_{play}}},\mathbf{x}_{\mathrm{hp_{hptf}}},\mathbf{x}_{\mathrm{m_{hptf}}})}.\end{aligned} \tag{C.40}$$

Equations 4.29 and C.35 allow to describe the equalization requirements for *blocked auditory canal recording* and human head playback in relation to *AH HPTFs* by

$$\begin{aligned}\mathbf{H}^{\mathrm{ind,b,h}}_{\mathrm{eq_m}}(\mathbf{x}_{\mathrm{hp_{play}}},\mathbf{x}_{\mathrm{m_{rec}}}) &= \frac{1}{\mathbf{H}^{\mathrm{ah,h}}_{\mathrm{hptf_{ahm}}}(\mathbf{x}_{\mathrm{hp_{hptf}}})} \cdot \frac{\mathbf{H}^{\mathrm{ah,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{hp_{hptf}}})}{\mathbf{H}^{\mathrm{ind,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{e}}}(\mathbf{x}_{\mathrm{hp_{play}}})} \cdot \\ &\cdot \frac{\mathbf{H}^{\mathrm{ind,h}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{e}}}(\mathbf{x}_{\mathrm{h_{ref}}},\mathbf{x}_{\mathrm{ls_{ref}}},\mathbf{x}_{\mathrm{hp_{ref}}})}{\mathbf{H}^{\mathrm{ind,b,h}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{m}}}(\mathbf{x}_{\mathrm{h_{ref}}},\mathbf{x}_{\mathrm{ls_{ref}}},\mathbf{x}_{\mathrm{m_{rec}}},\mathbf{x}_{\mathrm{hp_{rec}}})} \cdot \frac{\mathbf{H}_{\mathrm{ahm}} \cdot \mathbf{H}_{\mathrm{i_{ah}}}}{\mathbf{H}_{\mathrm{m}} \cdot \mathbf{H}_{\mathrm{i_m}}}.\end{aligned} \tag{C.41}$$

Equalization: Human Head Playback, Artificial Head Recording Equations C.22 and C.30 allow formulating the equalization requirements for BS with headphone reference scene, AH recording, and human head playback by

$$\begin{aligned}&\mathbf{H}^{\mathrm{ind,h}}_{\mathrm{eq_{ah}}}(\mathbf{x}_{\mathrm{hp_{play}}}) = \\ &= \frac{\mathbf{H}^{\mathrm{ind,h}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{e}}}(\mathbf{x}_{\mathrm{h_{ref}}},\mathbf{x}_{\mathrm{ls_{ref}}},\mathbf{x}_{\mathrm{hp_{ref}}})}{\mathbf{H}^{\mathrm{ah,h}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{h_{ref}}},\mathbf{x}_{\mathrm{ls_{ref}}},\mathbf{x}_{\mathrm{hp_{rec}}})} \cdot \frac{1}{\mathbf{H}_{\mathrm{ahm}} \cdot \mathbf{H}_{\mathrm{i_{ah}}} \cdot \mathbf{H}_{\mathrm{o_{hp}}} \cdot \mathbf{H}^{\mathrm{ind,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{e}}}(\mathbf{x}_{\mathrm{hp_{play}}})}.\end{aligned} \tag{C.42}$$

Equations 4.24 and C.42 yield the link between the equalization necessities for *AH recordings* played back to a human listener and *probe microphone HPTFs* by

$$\begin{aligned}\mathbf{H}^{\mathrm{ind,h}}_{\mathrm{eq_{ah}}}(\mathbf{x}_{\mathrm{hp_{play}}}) &= \frac{1}{\mathbf{H}^{\mathrm{ind,h}}_{\mathrm{hptf_{pm}}}(\mathbf{x}_{\mathrm{hp_{hptf}}},\mathbf{x}_{\mathrm{pm_{hptf}}})} \cdot \frac{\mathbf{H}^{\mathrm{ind,h}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{e}}}(\mathbf{x}_{\mathrm{h_{ref}}},\mathbf{x}_{\mathrm{ls_{ref}}},\mathbf{x}_{\mathrm{hp_{ref}}})}{\mathbf{H}^{\mathrm{ah,h}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{h_{ref}}},\mathbf{x}_{\mathrm{ls_{ref}}},\mathbf{x}_{\mathrm{hp_{rec}}})} \cdot \\ &\cdot \frac{\mathbf{H}^{\mathrm{ind,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{pm}}}(\mathbf{x}_{\mathrm{hp_{hptf}}},\mathbf{x}_{\mathrm{pm_{hptf}}})}{\mathbf{H}^{\mathrm{ind,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{e}}}(\mathbf{x}_{\mathrm{hp_{play}}})} \cdot \frac{\mathbf{H}_{\mathrm{pm}} \cdot \mathbf{H}_{\mathrm{i_{pm}}}}{\mathbf{H}_{\mathrm{ahm}} \cdot \mathbf{H}_{\mathrm{i_{ah}}}}.\end{aligned} \tag{C.43}$$

With equations 4.27 and C.42, the equalization requirements for the human head playback of *AH recordings* and *blocked auditory canal HPTFs* are computed to

$$\begin{aligned}\mathbf{H}^{\mathrm{ind,h}}_{\mathrm{eq_{ah}}}(\mathbf{x}_{\mathrm{hp_{play}}}) &= \frac{1}{\mathbf{H}^{\mathrm{ind,h,b}}_{\mathrm{hptf_m}}(\mathbf{x}_{\mathrm{hp_{hptf}}},\mathbf{x}_{\mathrm{m_{hptf}}})} \cdot \frac{1}{\mathbf{H}^{\mathrm{ind,h}}_{\mathbf{p}_{\mathrm{m_b}},\mathbf{p}_{\mathrm{e}},\mathrm{hp}}(\mathbf{x}_{\mathrm{hp_{play}}},\mathbf{x}_{\mathrm{hp_{hptf}}},\mathbf{x}_{\mathrm{m_{hptf}}})} \cdot \\ &\cdot \frac{\mathbf{H}^{\mathrm{ind,h}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{e}}}(\mathbf{x}_{\mathrm{h_{ref}}},\mathbf{x}_{\mathrm{ls_{ref}}},\mathbf{x}_{\mathrm{hp_{ref}}})}{\mathbf{H}^{\mathrm{ah,h}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{h_{ref}}},\mathbf{x}_{\mathrm{ls_{ref}}},\mathbf{x}_{\mathrm{hp_{ref}}})} \cdot \frac{\mathbf{H}_{\mathrm{m}} \cdot \mathbf{H}_{\mathrm{i_m}}}{\mathbf{H}_{\mathrm{ahm}} \cdot \mathbf{H}_{\mathrm{i_{ah}}}}.\end{aligned} \tag{C.44}$$

For gathering a relationship between the equalization necessary for *AH recordings* played back to a human listener and *AH HPTFs*, equations 4.29 and C.42 are combined to

$$\begin{aligned}
&\mathbf{H}^{\mathrm{ind,h}}_{\mathrm{eq_{ah}}}(\mathbf{x}_{\mathrm{hp_{play}}}) = \\
&= \frac{1}{\mathbf{H}^{\mathrm{ah,h}}_{\mathrm{hptf_{ahm}}}(\mathbf{x}_{\mathrm{hp_{hptf}}})} \cdot \frac{\mathbf{H}^{\mathrm{ah,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{hp_{hptf}}})}{\mathbf{H}^{\mathrm{ind,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{e}}}(\mathbf{x}_{\mathrm{hp_{play}}})} \cdot \frac{\mathbf{H}^{\mathrm{ind,h}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{e}}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{hp_{ref}}})}{\mathbf{H}^{\mathrm{ah,h}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{hp_{rec}}})}.
\end{aligned} \tag{C.45}$$

Equalization: Artificial Head Playback, Probe Microphone Recording The equalization requirements for the AH playback of probe microphone recordings are given combining equations C.24 and C.30 by

$$\begin{aligned}
&\mathbf{H}^{\mathrm{ah,h}}_{\mathrm{eq_{pm}}}(\mathbf{x}_{\mathrm{hp_{play}}}, \mathbf{x}_{\mathrm{pm_{rec}}}) = \\
&= \frac{1}{\mathbf{H}_{\mathrm{pm}} \cdot \mathbf{H}_{\mathrm{i_{pm}}} \cdot \mathbf{H}_{\mathrm{o_{hp}}} \cdot \mathbf{H}^{\mathrm{ah,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{hp_{play}}})} \cdot \frac{\mathbf{H}^{\mathrm{ah,h}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{hp_{ref}}})}{\mathbf{H}^{\mathrm{ind,h}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{pm}}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{pm_{rec}}}, \mathbf{x}_{\mathrm{hp_{rec}}})}.
\end{aligned} \tag{C.46}$$

With equations 4.24 and C.46, the equalization necessities for the AH playback of *probe microphone recordings* with regard to *probe microphone HPTFs* are given by

$$\begin{aligned}
\mathbf{H}^{\mathrm{ah,h}}_{\mathrm{eq_{pm}}}(\mathbf{x}_{\mathrm{hp_{play}}}, \mathbf{x}_{\mathrm{pm_{rec}}}) &= \frac{1}{\mathbf{H}^{\mathrm{ind,h}}_{\mathrm{hptf_{pm}}}(\mathbf{x}_{\mathrm{hp_{hptf}}}, \mathbf{x}_{\mathrm{pm_{hptf}}})} \cdot \\
&\cdot \frac{\mathbf{H}^{\mathrm{ind,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{pm}}}(\mathbf{x}_{\mathrm{hp_{hptf}}}, \mathbf{x}_{\mathrm{pm_{hptf}}})}{\mathbf{H}^{\mathrm{ah,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{hp_{play}}})} \cdot \frac{\mathbf{H}^{\mathrm{ah,h}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{hp_{ref}}})}{\mathbf{H}^{\mathrm{ind,h}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{pm}}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{pm_{rec}}}, \mathbf{x}_{\mathrm{hp_{rec}}})}.
\end{aligned} \tag{C.47}$$

To determine equalization requirements for the AH playback of *probe microphone recordings* dependent on *blocked auditory canal HPTFs*, equations 4.27 and C.46 are combined to

$$\begin{aligned}
\mathbf{H}^{\mathrm{ah,h}}_{\mathrm{eq_{pm}}}(\mathbf{x}_{\mathrm{hp_{play}}}, \mathbf{x}_{\mathrm{pm_{rec}}}) &= \frac{1}{\mathbf{H}^{\mathrm{ind,h,b}}_{\mathrm{hptf_{m}}}(\mathbf{x}_{\mathrm{hp_{hptf}}}, \mathbf{x}_{\mathrm{m_{hptf}}})} \frac{\mathbf{H}^{\mathrm{ind,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{e}}}(\mathbf{x}_{\mathrm{hp_{play}}})}{\mathbf{H}^{\mathrm{ah,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{hp_{play}}})} \frac{\mathbf{H}_{\mathrm{m}} \cdot \mathbf{H}_{\mathrm{i_{m}}}}{\mathbf{H}_{\mathrm{pm}} \cdot \mathbf{H}_{\mathrm{i_{pm}}}} \cdot \\
\cdot \frac{\mathbf{H}^{\mathrm{ah,h}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{hp_{ref}}})}{\mathbf{H}^{\mathrm{ind,h}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{pm}}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{pm_{rec}}}, \mathbf{x}_{\mathrm{hp_{rec}}})} &\cdot \frac{1}{\mathbf{H}^{\mathrm{ind,h}}_{\mathbf{p}_{\mathrm{m_b}},\mathbf{p}_{\mathrm{e}},\mathrm{hp}}(\mathbf{x}_{\mathrm{hp_{play}}}, \mathbf{x}_{\mathrm{hp_{hptf}}}, \mathbf{x}_{\mathrm{m_{hptf}}})}.
\end{aligned} \tag{C.48}$$

With equations 4.29 and C.46, equalization requirements for the AH playback of *probe microphone recordings* are given with regard to *AH HPTFs* by

$$\begin{aligned}
\mathbf{H}^{\mathrm{ah,h}}_{\mathrm{eq_{pm}}}(\mathbf{x}_{\mathrm{hp_{play}}}, \mathbf{x}_{\mathrm{pm_{rec}}}) &= \frac{1}{\mathbf{H}^{\mathrm{ah,h}}_{\mathrm{hptf_{ahm}}}(\mathbf{x}_{\mathrm{hp_{hptf}}})} \cdot \frac{\mathbf{H}^{\mathrm{ah,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{hp_{hptf}}})}{\mathbf{H}^{\mathrm{ah,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{hp_{play}}})} \cdot \\
&\cdot \frac{\mathbf{H}^{\mathrm{ah,h}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{hp_{ref}}})}{\mathbf{H}^{\mathrm{ind,h}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{pm}}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{pm_{rec}}}, \mathbf{x}_{\mathrm{hp_{rec}}})} \cdot \frac{\mathbf{H}_{\mathrm{ahm}} \cdot \mathbf{H}_{\mathrm{i_{ah}}}}{\mathbf{H}_{\mathrm{pm}} \cdot \mathbf{H}_{\mathrm{i_{pm}}}}.
\end{aligned} \tag{C.49}$$

Equalization: Artificial Head Playback, Blocked Auditory Canal Recording Equations C.26 and C.30 describe AH playback of blocked auditory canal recordings by

$$\begin{aligned}&\mathbf{H}^{\mathrm{ah,h}}_{\mathrm{eq_{ah}}}(\mathbf{x}_{\mathrm{hp_{play}}}, \mathbf{x}_{\mathrm{m_{rec}}}) = \\ &= \frac{\mathbf{H}^{\mathrm{ah,h}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{hp_{ref}}})}{\mathbf{H}^{\mathrm{ind,b,h}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{m}}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{m_{rec}}}, \mathbf{x}_{\mathrm{hp_{rec}}})} \cdot \frac{1}{\mathbf{H}_{\mathrm{m}} \cdot \mathbf{H}_{\mathrm{i_m}} \cdot \mathbf{H}_{\mathrm{o_{hp}}} \cdot \mathbf{H}^{\mathrm{ah,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{hp_{play}}})}.\end{aligned} \tag{C.50}$$

With equations 4.24 and C.50, the equalization for *blocked auditory canal recordings* and AH playback is formulated in dependence of *probe microphone HPTFs* by

$$\begin{aligned}\mathbf{H}^{\mathrm{ah,h}}_{\mathrm{eq_{ah}}}(\mathbf{x}_{\mathrm{hp_{play}}}, \mathbf{x}_{\mathrm{m_{rec}}}) = &\frac{1}{\mathbf{H}^{\mathrm{ind,h}}_{\mathrm{hptf_{pm}}}(\mathbf{x}_{\mathrm{hp_{hptf}}}, \mathbf{x}_{\mathrm{pm_{hptf}}})} \cdot \frac{\mathbf{H}^{\mathrm{ind,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{pm}}}(\mathbf{x}_{\mathrm{hp_{hptf}}}, \mathbf{x}_{\mathrm{pm_{hptf}}})}{\mathbf{H}^{\mathrm{ah,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{hp_{play}}})} \cdot \\ &\cdot \frac{\mathbf{H}^{\mathrm{ah,h}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{hp_{ref}}})}{\mathbf{H}^{\mathrm{ind,b,h}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{m}}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{m_{rec}}}, \mathbf{x}_{\mathrm{hp_{rec}}})} \cdot \frac{\mathbf{H}_{\mathrm{pm}} \cdot \mathbf{H}_{\mathrm{i_{pm}}}}{\mathbf{H}_{\mathrm{m}} \cdot \mathbf{H}_{\mathrm{i_m}}}.\end{aligned} \tag{C.51}$$

With equations 4.27 and C.50, the equalization for the AH playback of *blocked auditory canal recordings* is given using *blocked auditory canal HPTFs* by

$$\begin{aligned}\mathbf{H}^{\mathrm{ah,h}}_{\mathrm{eq_{ah}}}(\mathbf{x}_{\mathrm{hp_{play}}}, \mathbf{x}_{\mathrm{m_{rec}}}) = &\frac{1}{\mathbf{H}^{\mathrm{ind,h,b}}_{\mathrm{hptf_m}}(\mathbf{x}_{\mathrm{hp_{hptf}}}, \mathbf{x}_{\mathrm{m_{hptf}}})} \cdot \frac{\mathbf{H}^{\mathrm{ind,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{e}}}(\mathbf{x}_{\mathrm{hp_{play}}})}{\mathbf{H}^{\mathrm{ah,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{hp_{play}}})} \cdot \\ \cdot \frac{1}{\mathbf{H}^{\mathrm{ind,h}}_{\mathbf{p}_{\mathrm{m_b}},\mathbf{p}_{\mathrm{e}},\mathrm{hp}}(\mathbf{x}_{\mathrm{hp_{play}}}, \mathbf{x}_{\mathrm{hp_{hptf}}}, \mathbf{x}_{\mathrm{m_{hptf}}})} &\cdot \frac{\mathbf{H}^{\mathrm{ah,h}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{hp_{ref}}})}{\mathbf{H}^{\mathrm{ind,b,h}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{m}}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{m_{rec}}}, \mathbf{x}_{\mathrm{hp_{rec}}})}.\end{aligned} \tag{C.52}$$

Regarding an AH as a specific individual head, the equations

$$\mathbf{H}^{\mathrm{ind,h}}_{\mathbf{p}_{\mathrm{m_b}},\mathbf{p}_{\mathrm{e}},\mathrm{ls}}(\mathbf{x}_{\mathrm{m_{rec}}}, \mathbf{x}_{\mathrm{hp_{rec}}}, \mathbf{x}_{\mathrm{hp_{ref}}}) = \mathbf{H}^{\mathrm{ah,h}}_{\mathbf{p}_{\mathrm{m_b}},\mathbf{p}_{\mathrm{ahm_e}},\mathrm{ls}}(\mathbf{x}_{\mathrm{m_{rec}}}, \mathbf{x}_{\mathrm{hp_{rec}}}, \mathbf{x}_{\mathrm{hp_{ref}}}), \tag{C.53}$$

$$\mathbf{H}^{\mathrm{ind,b,h}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{m}}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{m_{rec}}}, \mathbf{x}_{\mathrm{hp_{rec}}}) = \mathbf{H}^{\mathrm{ah,b,h}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{m}}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{m_{rec}}}, \mathbf{x}_{\mathrm{hp_{rec}}}), \quad \text{and} \tag{C.54}$$

$$\mathbf{H}^{\mathrm{ind,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{e}}}(\mathbf{x}_{\mathrm{hp_{hptf}}}) = \mathbf{H}^{\mathrm{ah,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{hp_{hptf}}}) \tag{C.55}$$

hold (cf. section 4.6.2). Hence, equation C.52 can be simplified using equation 4.52 by

$$\mathbf{H}^{\mathrm{ah,h}}_{\mathrm{eq_{ah}}}(\mathbf{x}_{\mathrm{hp_{play}}}) = \frac{1}{\mathbf{H}^{\mathrm{ah,h,b}}_{\mathrm{hptf_{ahm}}}(\mathbf{x}_{\mathrm{hp_{hptf}}})} \cdot \frac{\mathbf{H}^{\mathrm{ah,h}}_{\mathbf{p}_{\mathrm{m_b}},\mathbf{p}_{\mathrm{ahm_e}},\mathrm{ls}}(\mathbf{x}_{\mathrm{m_{rec}}}, \mathbf{x}_{\mathrm{hp_{rec}}}, \mathbf{x}_{\mathrm{hp_{ref}}})}{\mathbf{H}^{\mathrm{ah,h}}_{\mathbf{p}_{\mathrm{m_b}},\mathbf{p}_{\mathrm{ahm_e}},\mathrm{hp}}(\mathbf{x}_{\mathrm{hp_{play}}}, \mathbf{x}_{\mathrm{hp_{hptf}}}, \mathbf{x}_{\mathrm{m_{hptf}}})}. \tag{C.56}$$

With equations 4.29 and C.50, the equalization requirements for *blocked auditory canal recording* and AH playback related to *AH HPTFs* are given by

$$\begin{aligned}\mathbf{H}^{\mathrm{ah,h}}_{\mathrm{eq_{ah}}}(\mathbf{x}_{\mathrm{hp_{play}}}, \mathbf{x}_{\mathrm{m_{rec}}}) = &\frac{1}{\mathbf{H}^{\mathrm{ah,h}}_{\mathrm{hptf_{ahm}}}(\mathbf{x}_{\mathrm{hp_{hptf}}})} \cdot \frac{\mathbf{H}^{\mathrm{ah,h}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{hp_{ref}}})}{\mathbf{H}^{\mathrm{ind,b,h}}_{u_{\mathrm{ls}},\mathbf{p}_{\mathrm{m}}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{m_{rec}}}, \mathbf{x}_{\mathrm{hp_{rec}}})} \cdot \\ &\cdot \frac{\mathbf{H}^{\mathrm{ah,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{hp_{hptf}}})}{\mathbf{H}^{\mathrm{ah,h}}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{ahm}}}(\mathbf{x}_{\mathrm{hp_{play}}})} \cdot \frac{\mathbf{H}_{\mathrm{ahm}} \cdot \mathbf{H}_{\mathrm{i_{ah}}}}{\mathbf{H}_{\mathrm{m}} \cdot \mathbf{H}_{\mathrm{i_m}}}.\end{aligned} \tag{C.57}$$

Equalization: Artificial Head Playback, Artificial Head Recording Equations C.27 and C.30 allow giving the equalization requirements for AH playback of AH recordings by

$$\mathbf{H}_{\mathrm{eq}_{\mathrm{ah}}}^{\mathrm{ah,h}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{play}}}) = \frac{1}{\mathbf{H}_{\mathrm{ahm}} \cdot \mathbf{H}_{\mathrm{i}_{\mathrm{ah}}} \cdot \mathbf{H}_{\mathrm{o}_{\mathrm{hp}}} \cdot \mathbf{H}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{ahm}}}^{\mathrm{ah,h}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{play}}})}. \tag{C.58}$$

With equations 4.24 and C.58, the equalization requirements for the AH playback of *AH recordings* are given dependent on *probe microphone HPTFs* by

$$\begin{aligned} &\mathbf{H}_{\mathrm{eq}_{\mathrm{ah}}}^{\mathrm{ah,h}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{play}}}) = \\ &= \frac{1}{\mathbf{H}_{\mathrm{hptf}_{\mathrm{pm}}}^{\mathrm{ind,h}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{hptf}}}, \mathbf{x}_{\mathrm{pm}_{\mathrm{hptf}}})} \cdot \frac{\mathbf{H}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{pm}}}^{\mathrm{ind,h}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{hptf}}}, \mathbf{x}_{\mathrm{pm}_{\mathrm{hptf}}})}{\mathbf{H}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{ahm}}}^{\mathrm{ah,h}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{play}}})} \cdot \frac{\mathbf{H}_{\mathrm{pm}} \cdot \mathbf{H}_{\mathrm{i}_{\mathrm{pm}}}}{\mathbf{H}_{\mathrm{ahm}} \cdot \mathbf{H}_{\mathrm{i}_{\mathrm{ah}}}}. \end{aligned} \tag{C.59}$$

The equalization requirements for the AH playback of *AH recordings* related to *blocked auditory canal HPTFs* are given using equations 4.27 and C.58 by

$$\begin{aligned} \mathbf{H}_{\mathrm{eq}_{\mathrm{ah}}}^{\mathrm{ah,h}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{play}}}) = &\frac{1}{\mathbf{H}_{\mathrm{hptf}_{\mathrm{m}}}^{\mathrm{ind,h,b}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{hptf}}}, \mathbf{x}_{\mathrm{m}_{\mathrm{hptf}}})} \cdot \frac{1}{\mathbf{H}_{\mathbf{p}_{\mathrm{m}_{\mathrm{b}}},\mathbf{p}_{\mathrm{e}},\mathrm{hp}}^{\mathrm{ind,h}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{play}}}, \mathbf{x}_{\mathrm{hp}_{\mathrm{hptf}}}, \mathbf{x}_{\mathrm{m}_{\mathrm{hptf}}})} \cdot \\ &\cdot \frac{\mathbf{H}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{e}}}^{\mathrm{ind,h}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{play}}})}{\mathbf{H}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{ahm}}}^{\mathrm{ah,h}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{play}}})} \cdot \frac{\mathbf{H}_{\mathrm{m}} \cdot \mathbf{H}_{\mathrm{i}_{\mathrm{m}}}}{\mathbf{H}_{\mathrm{ahm}} \cdot \mathbf{H}_{\mathrm{i}_{\mathrm{ah}}}}. \end{aligned} \tag{C.60}$$

With equations 4.29 and C.58, the equalization required for the AH playback of *AH recordings* is given related to *AH HPTFs* by

$$\mathbf{H}_{\mathrm{eq}_{\mathrm{ah}}}^{\mathrm{ah,h}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{play}}}) = \frac{1}{\mathbf{H}_{\mathrm{hptf}_{\mathrm{ahm}}}^{\mathrm{ah,h}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{hptf}}})} \cdot \frac{\mathbf{H}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{ahm}}}^{\mathrm{ah,h}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{hptf}}})}{\mathbf{H}_{\mathbf{u}_{\mathrm{hp}},\mathbf{p}_{\mathrm{ahm}}}^{\mathrm{ah,h}}(\mathbf{x}_{\mathrm{hp}_{\mathrm{play}}})}. \tag{C.61}$$

D Mathematical and System-Theoretic Derivations

D.1 Discrete Formulation of the Smoothing Windows

The temporal windows for auditory-adapted exponential transfer function smoothing (AAS) are defined in section 5.1.5 as real-valued, exponentially decaying temporal windows. These functions are formulated with the analysis frequency dependent time constants $a_\kappa > 0$ and the unit step function $\varepsilon(t)$ after Oppenheim et al. (1998, equation 1.70) by

$$w_{\mathrm{e}_\kappa}(t) = \mathrm{e}^{-a_\kappa t}\varepsilon(t) \quad \circ\!\!-\!\!\bullet \quad W_{\mathrm{e}_\kappa}(\omega) = \frac{1}{a_\kappa + \mathrm{j}\omega}. \tag{D.1}$$

For the digital AAS implementation, a discrete formulation is derived in the following. Yang et al. (2009, equation E 3.15.1, nomenclature adapted) give the infinite duration real-valued exponential sequence

$$s[n] = b^n \varepsilon[n], \quad \text{with } \left|b\right| < 1, \tag{D.2}$$

using the unit step sequence $\varepsilon[n]$ according to Oppenheim et al. (1998, equation 1.64). The discrete Fourier transform (DFT) corresponding to equation D.2 is given by Yang et al. (2009, equation E 3.15.4) by

$$S[k] = \frac{1 - b^N}{1 - b\,\mathrm{e}^{-\mathrm{j}\frac{2\pi k}{N}}}, \quad \text{with } k = 0, \ldots, N-1. \tag{D.3}$$

By substituting $b = \mathrm{e}^{-\frac{a_\kappa}{f_\mathrm{s}}}$, $\left|b\right| < 1$ holds true since $a_\kappa > 0$ represents positive time constants and f_s the positive sample rate. Therefore, equations D.2 and D.3 can be used to derive the DFT pair corresponding to equation D.1 to

$$w_{\mathrm{e}_\kappa}[n] = \mathrm{e}^{-\frac{n a_\kappa}{f_\mathrm{s}}}\varepsilon[n] \quad \circ\!\!-\!\!\bullet \quad W_{\mathrm{e}_\kappa}[k] = \frac{1 - \mathrm{e}^{-\frac{N a_\kappa}{f_\mathrm{s}}}}{1 - \mathrm{e}^{-\mathrm{j}\frac{2\pi k}{N} - \frac{a_\kappa}{f_\mathrm{s}}}}. \tag{D.4}$$

D.2 Extreme Values of the Auditory-Adapted Analysis Windows

The indices $n_{\mathrm{max},\eta_\kappa}$ of the samples with maximum values of the discrete auditory-adapted analysis (AAA) windows given by equation 2.22 are computed based on their derivatives

$$\begin{aligned}\frac{d}{dn} w_{\eta_\kappa}[n] &= \frac{2a_\kappa}{(\eta-1)!}\frac{d}{dn}\left[(a_\kappa n/f_\mathrm{s})^{\eta-1}\,\mathrm{e}^{-a_\kappa n/f_\mathrm{s}}\right] \\ &= \frac{2a_\kappa^2}{f_\mathrm{s}(\eta-1)!}(a_\kappa n/f_\mathrm{s})^{\eta-2}\,\mathrm{e}^{-a_\kappa n/f_\mathrm{s}}\Big[(\eta-1) - a_\kappa n/f_\mathrm{s}\Big], \quad n \geq 0.\end{aligned} \tag{D.5}$$

In the case discussed here, window functions with $\eta > 1$ are of interest. Extreme values occur at zeros of the derivate, that is for

$$\frac{d}{dn} w_{\eta_\kappa}\left[n_{\text{ext},\eta_\kappa}\right] \overset{!}{=} 0 \quad \Rightarrow \quad \begin{cases} \text{I.)} & n_{\text{ext},\eta_\kappa} = 0 \\ \text{II.)} & \eta - 1 = a_\kappa n_{\text{ext},\eta_\kappa} / f_\text{s} \end{cases}. \tag{D.6}$$

Taking into account that $w_{\eta_\kappa}[0] = 0$ holds for $\eta > 1$, maxima occur at

$$n_{\text{max},\eta_\kappa} = f_\text{s}\left(\eta - 1\right) / a_\kappa, \quad \forall \eta > 1. \tag{D.7}$$

With equations 2.26 and 2.27, equation D.7 can be rewritten by

$$n_{\text{max},\eta_\kappa} = \frac{f_\text{s}\left(\eta - 1\right)}{\pi c \Delta f_\text{GV}\left(f_{\text{A}_\kappa}\right)} \sqrt{2^{1/\eta} - 1}, \quad \forall \eta > 1. \tag{D.8}$$

The corresponding maximum values are given by

$$w_{\eta_\kappa}\left[n_{\text{max},\eta_\kappa}\right] = \frac{2a_\kappa}{(\eta - 1)!}\left(\eta - 1\right)^{\eta - 1} \mathrm{e}^{1-\eta}, \quad \forall \eta > 1. \tag{D.9}$$

The sample index n_{α,η_κ} at the window attenuation $\alpha = 10^{\frac{\Delta L}{20}}$ with respect to the window maximum is given with $\Delta L = L_\alpha - L_\text{max}$ based on

$$\begin{aligned} w_{\eta_\kappa}\left[n_{\alpha,\eta_\kappa}\right] &= \frac{2a_\kappa}{(\eta - 1)!}\left(a_\kappa n_{\alpha,\eta_\kappa} / f_\text{s}\right)^{\eta - 1} \mathrm{e}^{-a_\kappa n_{\alpha,\eta_\kappa} / f_\text{s}} \\ &\overset{!}{=} \alpha \frac{2a_\kappa}{(\eta - 1)!}\left(\eta - 1\right)^{\eta - 1} \mathrm{e}^{1-\eta}, \quad \forall \eta > 1 \end{aligned} \tag{D.10}$$

$$\Rightarrow -\frac{a_\kappa n_{\alpha,\eta_\kappa}}{f_\text{s}\left(\eta - 1\right)} \mathrm{e}^{-\frac{a_\kappa n_{\alpha,\eta_\kappa}}{f_\text{s}(\eta - 1)}} \overset{!}{=} -\frac{1}{\mathrm{e}} \alpha^{\frac{1}{\eta - 1}}, \quad \forall \eta > 1. \tag{D.11}$$

Substituting $x = -\frac{a_\kappa n_{\alpha,\eta_\kappa}}{f_\text{s}(\eta - 1)}$ and therefore $n_{\alpha,\eta_\kappa} = -\frac{x f_\text{s}(\eta - 1)}{a_\kappa}$, the simplification

$$x\,\mathrm{e}^x = -\frac{1}{\mathrm{e}} \alpha^{\frac{1}{\eta - 1}} \tag{D.12}$$

is possible. Using the Lambert W-function (Corless et al. 1996), enabling the computation of x for equations of the form $y = x\,\mathrm{e}^x$, n_{α,η_κ} is given using equations 2.26 and 2.27 by

$$n_{\alpha,\eta_\kappa} = -f_\text{s} \frac{x\left(\eta - 1\right)\sqrt{2^{1/\eta} - 1}}{\pi c \Delta f_\text{GV}\left(f_{\text{A}_\kappa}\right)}, \quad \forall \eta > 1. \tag{D.13}$$

This sample index corresponds to the ΔL decay times of the AAA windows. In the context of this thesis, it is used to correct the minimum temporal resolution displayed in AAA spectrograms by the 60 dB decay times of the AAA windows (cf. e. g. figure 2.6).

Bibliography

Aarts R. M.: Enlarging the Sweet-Spot for Stereophony by Time/Intensity Trading. In *94th AES Convention* (1993) (Preprint 3473)

Abramowitz M., I. A. Stegun: *Handbook of Mathematical Functions.* 10th Edition (United States of America, Department of Commerce, National Bureau of Standards, 1972) (Applied Mathematics Series 55)

Adams G.: Time Dependence of Loudspeaker Power Output in Small Rooms. J. Audio Eng. Soc. **37**, 203–209 (1989)

Affel H. A., R. W. Chesnut, R. H. Mills: Symposium on Wire Transmission of Symphonic Music and Its Reproduction in Auditory Perspective – Transmission Lines. Bell System Technical Journal **13**, 285–300 (1934)

Ahrens J., S. Spors: Rendering of virtual sound sources with arbitrary directivity in higher order Ambisonics. In *123rd AES Convention* (2007) (Convention Paper 7243)

Ahrens J., S. Spors: An Analytical Approach to Sound Field Reproduction Using Circular and Spherical Loudspeaker Distributions. Acta Acustica united with Acustica **94**, 988–999 (2008a)

Ahrens J., S. Spors: Analytical driving functions for higher order Ambisonics. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 373–376 (2008b)

Ahrens J., S. Spors: Focusing of virtual sound sources in higher order Ambisonics. In *124th AES Convention* (2008c) (Convention Paper 7378)

Ahrens J., S. Spors: Reproduction of a plane-wave sound field using planar and linear arrays of loudspeakers. In *3rd International Symposium on Communications, Control and Signal Processing (ISCCSP)*, 1486–1491 (2008d)

Akeroyd M. A., J. Chambers, D. Bullock, A. R. Palmer, A. Q. Summerfield, P. A. Nelson, S. Gatehouse: The binaural performance of a cross-talk cancellation system with matched or mismatched setup and playback acoustics. J. Acoust. Soc. Am. **121**, 1056–1069 (2007)

Algazi V. R., C. Avendano, R. O. Duda: Elevation localization and head-related transfer function analysis at low frequencies. J. Acoust. Soc. Am. **109**, 1110–1122 (2001)

Allen R. G., E. L. Torick, B. B. Bauer: A Dynamic Presence Equalizer. In *37th AES Convention* (1969) (Preprint 702)

Aoki S., H. Miyata, K. Sugiyama: Stereo Reproduction with Good Localization over a Wide Listening Area. J. Audio Eng. Soc. **38**, 433–439 (1990)

Apple Inc.: *Core Audio Overview.* 2007-01-08 Edition. 1 Infinite Loop, Cupertino, CA 95014, USA (August 2007)

Asano F., Y. Suzuki, T. Sone: Role of spectral cues in median plane localization. J. Acoust. Soc. Am. **88**, 159–168 (1990)

Ashley J. R.: On the Psycho-Acoustic Basis for Two and Four Channel Home Music Systems. In *53rd AES Convention* (1976) (Preprint B-6)

Atal B. S., M. R. Schroeder, G. M. Sessler, J. E. West: Evaluation of Acoustic Properties of Enclosures by Means of Digital Computers. J. Acoust. Soc. Am. **40**, 428–433 (1966)

Aures W.: Der Wohlklang: Eine Funktion von Schärfe, Rauhigkeit, Klanghaftigkeit und Lautheit (Sensory pleasantness: A function of sharpness, roughness, tonality, and loudness). In *Fortschritte der Akustik, DAGA '84*, 735–738 (DPG, Bad Honnef, 1984)

Azzali A., A. Farina, G. Rovai, G. Boreanaz, G. Irato: Construction of a Car Stereo Audio Quality Index. In *117th AES Convention* (2004) (Convention Paper 6306)

Bai M. R., C.-C. Lee: Comparative Study of Design and Implementation Strategies of Automotive Virtual Surround Audio Systems. J. Audio Eng. Soc. **58**, 141–159 (2010)

Bai M. R., C.-W. Tung, C.-C. Lee: Optimal design of loudspeaker arrays for robust cross-talk cancellation using the Taguchi method and the genetic algorithm. J. Acoust. Soc. Am. **117**, 2802–2813 (2005)

Barbour J. L.: Elevation Perception: Phantom Images in the Vertical Hemi-Sphere. In *24th International AES Conference* (2003) (Paper Number 14)

Bass H. E., L. C. Sutherland, A. J. Zuckerwar: Atmospheric absorption of sound: Update. J. Acoust. Soc. Am. **88**, 2019–2021 (1990)

Bass H. E., L. C. Sutherland, A. J. Zuckerwar, D. T. Blackstock, D. M. Hester: Atmospheric absorption of sound: Further developments. J. Acoust. Soc. Am. **97**, 680–683 (1995)

Batke J.-M., F. Keiler: Robust Spatial Panning Functions for Nonuniform Loudspeaker Layouts. In *Fortschritte der Akustik, DAGA 2010*, 1061–1062 (Dt. Gesell. für Akustik e. V., Berlin, 2010)

Bauck J., D. H. Cooper: Generalized Transaural Stereo and Applications. J. Audio Eng. Soc. **44**, 683–705 (1996)

Bauer B. B.: Broadening the Area of Stereophonic Perception. J. Audio Eng. Soc. **8**, 91–94 (1960)

Baumgarte F., C. Faller: Binaural Cue Coding – Part I: Psychoacoustic Fundamentals and Design Principles. IEEE Transactions on Speech and Audio Processing **11**, 509–519 (2003)

Bech S.: The Influence of Stereophonic Width on the Perceived Quality of an Audiovisual Presentation Using a Multichannel Sound System. J. Audio Eng. Soc. **46**, 314–322 (1998)

Bedell E. H., I. Kerney: Symposium on Wire Transmission of Symphonic Music and Its Reproduction in Auditory Perspective – System Adaptation. Bell System Technical Journal **13**, 301–308 (1934)

Beerends J. G., F. E. de Caluwe: The Influence of Video Quality on Perceived Audio Quality and Vice Versa. J. Audio Eng. Soc. **47**, 355–362 (1999)

Begault D. R.: Virtual Acoustics, Aeronautics, and Communications. J. Audio Eng. Soc. **46**, 520–530 (1998)

Begault D. R.: Virtual Acoustic Displays for Teleconferencing: Intelligibility Advantage for "Telephone-Grade" Audio. J. Audio Eng. Soc. **47**, 824–828 (1999)

Begault D. R., M. R. Anderson, B. U. McClain: Spatially Modulated Auditory Alerts for Aviation. J. Audio Eng. Soc. **54**, 276–282 (2006)

Begault D. R., E. M. Wenzel, M. R. Anderson: Direct Comparison of the Impact of Head Tracking, Reverberation, and Individualized Head-Related Transfer Functions on the Spatial Perception of a Virtual Speech Source. J. Audio Eng. Soc. **49**, 904–916 (2001)

Benjamin E.: Signal Characteristics of Matrix and Discrete Multichannel Recordings. In *105th AES Convention* (1998) (Preprint 4859)

Benjamin E.: An Experimental Verification of Localization in Two-Channel Stereo. In *121st AES Convention* (2006) (Convention Paper 6968)

Beranek L. L.: *Acoustic Measurements.* (John Wiley & Sons, Inc., New York, 1949)

Berkhout A. J.: A Holographic Approach to Acoustic Control. J. Audio Eng. Soc. **36**, 977–995 (1988)

Berkhout A. J., D. de Vries: Acoustic holography for sound control. In *86th AES Convention* (1989) (Preprint 2801)

Berkhout A. J., D. de Vries, P. Vogel: Acoustic control by wave field synthesis. J. Acoust. Soc. Am. **93**, 2764–2778 (1993)

Bernfeld B.: Simple Equations For Multichannel Stereophonic Sound Localization. J. Audio Eng. Soc. **23**, 553–557 (1975)

Bixler O.: A practical binaural recording system. Transactions of the IRE Professional Group on Audio **1**, 14–22 (1953)

Blauert J.: Ein neuartiges Präsenzfilter (A novel presence filter). Fernseh- und Kino-Technik **3**, 75–78 (1970)

Blauert J.: *Spatial Hearing – The Psychophysics of Human Sound Localization.* Revised Edition (The MIT Press, Cambridge, Massachusetts, London, 1997)

Blauert J., J. Braasch: *Chapter* Räumliches Hören (Spatial hearing). In *Handbuch der Audiotechnik (Handbook of audio technology)*, ed. by S. Weinzierl (Springer, 2007)

Blauert J., U. Jekosch: Sound-Quality Evaluation – A Multi-Layered Problem. Acustica – Acta Acustica **83**, 747–753 (1997)

Blauert J., P. Tritthart: Ausnutzung von Verdeckungseffekten bei der Sprachkodierung (Exploiting masking effects for speech coding). In *Fortschritte der Akustik, DAGA '75*, 377–380 (Physik, Weinheim, 1975)

Blauert J., H. Lehnert, J. Sahrhage, H. Strauss: An Interactive Virtual-Environment Generator for Psychoacoustic Research. I: Architecture and Implementation. Acustica – Acta Acustica **86**, 94–102 (2000)

Blauert J., J. Braasch, J. Buchholz, H. S. Colburn, U. Jekosch, A. Kohlrausch, J. Mouriopoulos, V. Pulkki, A. Raake: Aural assessment by means of binaural algorithms – The AABBA project. In *International Symposium on Auditory and Audiological Research (ISAAR)* (2009)

Blauert J., V. Mellert, H.-J. Platte, P. Laws, H. Hudde, P. Scherer, T. Poulsen, D. Gottlob, G. Plenge: Wissenschaftliche Grundlagen der kopfbezogenen Stereophonie (Scientific basics of the head-related stereophony). Rundfunktechnische Mitteilungen **22**, 195–218 (1978)

Bocker P., H. Mrass: Zur Bestimmung des Freifeld-Übertragungsmaßes von Kopfhörern (On the determination of the free-field transfer function of headphones). Acustica **9**, 340–344 (1959)

Bodden M.: Modeling human sound-source localization and the cocktail-party-effect. Acta Acustica **1**, 43–55 (1993)

Boone M. M., W. P. J. de Bruijn: Improving speech intelligibility in teleconferencing by using Wave Field Synthesis. In *114th AES Convention* (2003) (Convention Paper 5800)

Boone M. M., E. N. G. Verheijen, P. F. van Tol: Spatial Sound-Field Reproduction by Wave-Field Synthesis. J. Audio Eng. Soc. **43**, 1003–1012 (1995)

Bortz J.: *Statistik für Human- und Sozialwissenschaftler (Statistics for human and social scientists).* 6th Edition (Springer Medizin Verlag, Heidelberg, 2005)

Brandenburg K.: Evaluation of quality for audio encoding at low bit rates. In *82nd AES Convention* (1987) (Preprint 2433)

Breebaart J., F. Nater, A. Kohlrausch: Spectral and Spatial Parameter Resolution Requirements for Parametric, Filter-Bank-Based HRTF Processing. J. Audio Eng. Soc. **58**, 126–140 (2010)

Bronkhorst A. W.: Localization of real and virtual sound sources. J. Acoust. Soc. Am. **98**, 2542–2553 (1995)

Bronkhorst A. W., T. Houtgast: Auditory distance perception in rooms. Nature **397**, 517–520 (1999)

Bronstein I. N., K. A. Semendjajew, G. Musiol, H. Mühlig: *Taschenbuch der Mathematik (Handbook of mathematics).* 5th Edition (Verlag Harri Deutsch, Frankfurt am Main, 2001)

Brungart D. S.: Auditory localization of nearby sources. III. Stimulus effects. J. Acoust. Soc. Am. **106**, 3589–3602 (1999)

Brungart D. S., W. M. Rabinowitz: Auditory localization of nearby sources. Head-related transfer functions. J. Acoust. Soc. Am. **106**, 1465–1479 (1999)

Brungart D. S., N. I. Durlach, W. M. Rabinowitz: Auditory localization of nearby sources. II. Localization of a broadband source. J. Acoust. Soc. Am. **106**, 1956–1968 (1999)

Brungart D. S., B. D. Simpson, A. J. Kordik: The Detectability of Headtracker Latency in Virtual Audio Displays. In *11th International Conference on Auditory Display (ICAD)*, 37–42 (2005)

Burkhard M. D., R. M. Sachs: Anthropometric manikin for acoustic research. J. Acoust. Soc. Am. **58**, 214–222 (1975)

Carlile S., D. Pralong: The location-dependent nature of perceptually salient features of the human head-related transfer functions. J. Acoust. Soc. Am. **95**, 3445–3459 (1994)

Chalupper J., H. Fastl: Dynamic Loudness Model (DLM) for Normal and Hearing-Impaired Listeners. Acta Acustica united with Acustica **88**, 378–386 (2002)

Chan J. C. K., C. D. Geisler: Estimation of eardrum acoustic pressure and of ear canal length from remote points in the canal. J. Acoust. Soc. Am. **87**, 1237–1247 (1990)

Chandler D. W., D. W. Grantham: Minimum audible movement angle in the horizontal plane as a function of stimulus frequency and bandwidth, source azimuth, and velocity. J. Acoust. Soc. Am. **91**, 1624–1636 (1992)

Christensen F., H. Møller, P. Minnaar, J. Plogsties, S. K. Olesen: Interpolating between Head-Related Transfer Functions Measured with Low Directional Resolution. In *107th AES Convention* (1999) (Preprint 5047)

Clapp S., A. Guthrie, J. Braasch, N. Xiang: Investigating Room Acoustics Using a 16-Channel, 2nd-Order Ambisonic Microphone. In *20th International Congress on Acoustics (ICA)* (2010)

Cohen M., O. N. N. Fernando, T. Nagai, K. Shimizu: “Back-Seat Driver”: Spatial Sound for Vehicular Way-Finding and Situation Awareness. In *IEEE Japan-China Joint Workshop on Frontier of Computer Science and Technology (FCST ’06)*, 109–115 (2006)

Cooper D. H., J. L. Bauck: Prospects for Transaural Recording. J. Audio Eng. Soc. **37**, 3–19 (1989)

Corless R. M., G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, D. E. Knuth: On the Lambert W function. Advances in Computational Mathematics **5**, 329–359 (1996)

Couling J.: Dolby Digital Surround Systems. In *14th AES UK Conference*, 156–160 (1999) (Paper Number ASC-19)

Cox T. J., P. D’Antonio: Room Optimizer: A Computer Program to Optimize the Placement of Listener, Loudspeakers, Acoustical Surface Treatment and Room Dimensions in Critical Listening Rooms. In *103rd AES Convention* (1997) (Preprint 4555)

Daniel J., R. Nicol, S. Moreau: Further Investigations of High Order Ambisonics and Wavefield Synthesis for Holophonic Sound Imaging. In *114th AES Convention* (2003) (Convention Paper 5788)

Daniel P., H. Fastl, T. Fedtke, K. Genuit, H.-P. Grabsch, T. Niederdränk, A. Schmitz, M. Vorländer, M. Zollner: Kunstkopftechnik – Eine Bestandsaufnahme (Artificial head technology – A survey). Nuntius Acusticus **6**, 1–58 (2007)

Davis M. F.: Loudspeaker Systems with Optimized Wide-Listening-Area Imaging. J. Audio Eng. Soc. **35**, 888–896 (1987)

Davis M. F., M. C. Fellers: Virtual Surround Presentation of Dolby AC-3 and Pro Logic Signals. In *103rd AES Convention* (1997) (Preprint 4542)

de Bruijn W., M. M. Boone, D. de Vries: Sound Localisation in a Videoconferencing System based on Wave Field Synthesis. In *108th AES Convention* (2000) (Preprint 5144)

de Vries D.: Sound Reinforcement by Wavefield Synthesis: Adaptation of the Synthesis Operator to the Loudspeaker Directivity Characteristics. J. Audio Eng. Soc. **44**, 1120–1131 (1996)

Dickreiter M.: *Mikrofon-Aufnahmetechnik (Microphone recording technique).* 3rd Edition (Hirzel, Stuttgart, 2003)

DIN 45 631: *Berechnung des Lautstärkepegels und der Lautheit aus dem Geräuschspektrum. Verfahren nach E. Zwicker (Calculation of loudness level and loudness from the sound spectrum. Zwicker method).* (Deutsche Norm, Beuth Verlag, Berlin, 1991)

DIN 45 631/A1: *Berechnung des Lautstärkepegels und der Lautheit aus dem Geräuschspektrum – Verfahren nach E. Zwicker – Änderung 1: Berechnung der Lautheit zeitvarianter Geräusche; mit CD-ROM (Calculation of loudness level and loudness from the sound spectrum – Zwicker method – Amendment 1: Calculation of the loudness of time-variant sound; with CD-ROM).* (Deutsche Norm, Beuth Verlag, Berlin, 2010)

DIN EN ISO 3382: *Akustik – Messung der Nachhallzeit von Räumen mit Hinweis auf andere akustische Parameter (Acoustics – Measurement of the reverberation time of rooms with reference to other acoustical parameters).* (Deutsche Norm, Beuth Verlag, Berlin, 2000)

DIN ISO 226: *Akustik – Normalkurven gleicher Lautstärkepegel (Acoustics – Normal equal-loudness-level contours).* (Deutsche Norm, Beuth Verlag, Berlin, 2006)

Djelani T., C. Pörschmann, J. Sahrhage, J. Blauert: An Interactive Virtual-Environment Generator for Psychoacoustic Research II: Collection of Head-Related Impulse Responses and Evaluation of Auditory Localization. Acustica – Acta Acustica **86**, 1046–1053 (2000)

Domnitz R.: The interaural time jnd as a simultaneous function of interaural time and interaural amplitude. J. Acoust. Soc. Am. **53**, 1549–1552 (1973)

Duda R. O., W. L. Martens: Range dependence of the response of a spherical head model. J. Acoust. Soc. Am. **104**, 3048–3058 (1998)

Dyreby J., S. Choisel: Equalization of loudspeaker resonances using second-order filters based on spatially distributed impulse response measurements. In *123rd AES Convention* (2007) (Convention Paper 7205)

Eargle J.: *Handbook of Recording Engineering.* 4th Edition (Birkhäuser, 2005)

Edwards A. S.: Accuracy of Auditory Depth Perception. J. Gen. Psychol. **52**, 327–329 (1955)

Ehret A., A. Gröschel, H. Purnhagen, J. Rödén: Coding of “2+2+2” surround sound content using the MPEG Surround standard. In *122nd AES Convention* (2007) (Convention Paper 6992)

Engdegård J., B. Resch, C. Falch, O. Hellmuth, J. Hilpert, A. Hoelzer, L. Terentiev, J. Breebaart, J. Koppens, E. Schuijers, W. Oomen: Spatial Audio Object Coding (SAOC) – The Upcoming MPEG Standard on Parametric Object Based Audio Coding. In *124th AES Convention* (2008) (Convention Paper 7377)

Enzner G., M. Krawczyk, F.-M. Hoffmann, M. Weinert: 3D Reconstruction of HRTF-Fields from 1D Continuous Measurements. In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 157–160 (2011)

Epp B., J. L. Verhey, M. Mauermann: Modeling cochlear dynamics: Interrelation between cochlea mechanics and psychoacoustics. J. Acoust. Soc. Am. **128**, 1870–1883 (2010)

Evans L. B., H. E. Bass, L. C. Sutherland: Atmospheric Absorption of Sound: Theoretical Predictions. J. Acoust. Soc. Am. **51**, 1565–1575 (1972)

Faller C.: Binaural Cue Coding: Rendering of Sources Mixed into a Mono Signal. In *Fortschritte der Akustik, DAGA '03*, 858–859 (Dt. Gesell. für Akustik e. V., Oldenburg, 2003)

Faller C.: Multiple-Loudspeaker Playback of Stereo Signals. J. Audio Eng. Soc. **54**, 1051–1064 (2006)

Faller C.: Parametric Coding of Spatial Audio. In *7th International Conference on Digital Audio Effects (DAFX '04)* (2004)

Farina A.: Simultaneous Measurement of Impulse Response and Distortion with a Swept-Sine Technique. In *108th AES Convention* (2000) (Preprint 5093)

Farina A., E. Ugolotti: Subjective comparison of different car audio systems by the auralization technique. In *103rd AES Convention* (1997) (Preprint 4587)

Farina A., E. Ugolotti: Numerical model of the sound field inside cars for the creation of virtual audible reconstructions. In 1st *International Conference on Digital Audio Effects (DAFX '98)* (1998)

Farina A., E. Ugolotti: Subjective Comparison between Stereo Dipole and 3D Ambisonic Surround Systems for Automotive Applications. In *16th International AES Conference* (1999) (Paper Number 16-048)

Fastl H.: Loudness and Masking Patterns of Narrow Noise Bands. Acustica **33**, 266–271 (1975)

Fastl H.: Temporal Masking Effects: II. Critical Band Noise Masker. Acustica **36**, 317–331 (1976/77)

Fastl H.: *Beschreibung dynamischer Hörempfindungen anhand von Mithörschwellen-Mustern (Dynamic hearing sensations and masking patterns).* (Hochschul, Freiburg, 1982)

Fastl H.: Gibt es *den* Frequenzgang von Kopfhörern? (Does *the* frequency response of headphones exist?). In *NTG-Fachberichte 91, Hörrundfunk 7*, 274–281 (VDE-Verlag, Berlin, 1986)

Fastl H.: Sound measurements based on features of the human hearing system. J. Acoust. Soc. Jpn. (E) **21**, 333–336 (2000)

Fastl H.: Neutralizing the meaning of sound for sound quality evaluations. In *17th International Congress on Acoustics (ICA)* (2001)

Fastl H.: Audio-visual interactions in loudness evaluation. In *18th International Congress on Acoustics (ICA)*, II 1161–1166 (2004)

Fastl H.: Psychoacoustics, sound quality and music. In *Inter-Noise 2007* (2007)

Fastl H.: Praktische Anwendungen der Psychoakustik (Practical applications of Psychoacoustics). In *Fortschritte der Akustik, DAGA 2010*, 5–10 (Dt. Gesell. für Akustik e. V., Berlin, 2010)

Fastl H., M. Bechly: Post masking with two maskers: Effects of bandwidth. J. Acoust. Soc. Am. **69**, 1753–1757 (1981)

Fastl H., E. Schorer: *Chapter* Critical bandwidth at low frequencies reconsidered. In *Auditory Frequency Selectivity*, ed. by B. C. J. Moore, R. D. Patterson, 311–318 (Plenum Press, New York and London, 1986)

Fastl H., E. Zwicker: A free-field equalizer for TDH 39 earphones. J. Acoust. Soc. Am. **73**, 312–314 (1983)

Fastl H., E. Zwicker: *Psychoacoustics – Facts and Models.* 3rd Edition (Springer, Berlin, Heidelberg, 2007)

Fastl H., D. Menzel, W. Maier: Entwicklung und Verifikation eines Lautheits-Thermometers (Development and verification of a loudness-thermometer). In *Fortschritte der Akustik, DAGA '06*, 669–670 (Dt. Gesell. für Akustik e. V., Berlin, 2006)

Fastl H., F. Völk, M. Straubinger: Standards for calculating loudness of stationary or time-varying sounds. In *Inter-Noise 2009* (2009)

Fastl H., W. Schmid, G. Theile, E. Zwicker: Schallpegel im Gehörgang für gleichlaute Schalle aus Kopfhörern oder Lautsprechern (Sound level in the ear canal for equally loud sounds from headphones versus loudspeakers). In *Fortschritte der Akustik, DAGA '85*, 471–474 (DPG, Bad Honnef, 1985)

Fastl H., H. Oberdanner, W. Schmid, I. Stemplinger, I. Hochmair-Desoyer, E. Hochmair: Zum Sprachverständnis von Cochlea-Implantat-Patienten bei Störgeräuschen (Speech intelligibility in noise for cochlea-implant patients). In *Fortschritte der Akustik, DAGA '98*, 358–359 (Dt. Gesell. für Akustik e. V., Oldenburg, 1998)

Fechner G. T.: *Elemente der Psychophysik (Elements of Psychophysics).* (Breitkopf & Härtel, Leipzig, 1860)

Felderhoff U., K. Reichenauer, G. Theile: Stabilität der Lokalisation bei verfälschter Reproduktion verschiedener Merkmale der binauralen Signale (Stability of Localization vs Distorted Reproduction of Binaural Cues). In *20. Tonmeistertagung, VDT International Convention*, 229–237 (Verband deutscher Tonmeister, VDT, 1998)

Feldtkeller R., E. Zwicker: *Das Ohr als Nachrichtenempfänger (The ear as a communication receiver).* (S. Hirzel Verlag, Stuttgart, 1956)

Fincham L. R.: Refinements in the Impulse Testing of Loudspeakers. J. Audio Eng. Soc. **33**, 133–140 (1985)

Fletcher H.: Symposium on Wire Transmission of Symphonic Music and Its Reproduction in Auditory Perspective – Basic Requirements. Bell System Technical Journal **13**, 239–244 (1934)

Fletcher H.: The Stereophonic Sound Film System – General Theory. J. Acoust. Soc. Am. **13**, 89–99 (1941)

Fletcher H., W. A. Munson: Loudness, Its Definition, Measurement and Calculation. J. Acoust. Soc. Am. **5**, 82–108 (1933)

Fletcher H., W. A. Munson: Relation Between Loudness and Masking. J. Acoust. Soc. Am. **9**, 1–10 (1937)

Fletcher H., J. C. Steinberg: The Dependence of the Loudness of a Complex Sound upon the Energy in the Various Frequency Regions of the Sound. Phys. Rev. **24**, 306–317 (1924)

Foster S.: Real-time implementation of spatial auditory displays. J. Acoust. Soc. Am. **92**, 2332 (1992)

Funasaka E., H. Suzuki: DVD-Audio Format. In *103rd AES Convention* (1997) (Preprint 4566)

Furuya T., K. Ozawa, Y. Suzuki: Two-dimensional localization of a phantom sound image controlled by the level differences among four loudspeakers in a vertical plane facing a listener. Acoust. Sci. & Tech. **25**, 493–495 (2004)

Gallo M., J. Anthonis, H. v. d. Auweraer, K. Janssens, L. de Oliveira: Evaluation of an Active Sound Quality Control System in a Virtual Car Driving Simulator. In *Inter-Noise 2010* (2010)

Gardner W. G.: Efficient Convolution without Input-Output Delay. J. Audio Eng. Soc. **43**, 127–136 (1995)

Gardner W. G.: *3-D Audio Using Loudspeakers*, PhD thesis, Massachusetts Inst. of Techn. (1997)

Gässler G.: Über die Hörschwelle für Schallereignisse mit verschieden breitem Frequenzspektrum (On the threshold in quiet of sounds with frequency spectra of different width). Acustica **4**, 408–414 (1954)

Gelfand S. A.: *Hearing. An introduction to psychological and physiological acoustics.* 4th Edition (Marcel Dekker, New York, 2004)

Genuit K.: Untersuchungen der psychoakustischen Eigenschaften von Hörereignissen bei der Kopfhörerwiedergabe (Investigations of psychoacoustic properties of auditory events in headphone playback). In *Fortschritte der Akustik, DAGA '86*, 489–492 (DPG, Bad Honnef, 1986)

Genuit K., A. Fiebig: Do we need new artificial heads? In *19th International Congress on Acoustics (ICA)* (2007)

Gerzon M. A.: Periphony: With-Height Sound Reproduction. J. Audio Eng. Soc. **21**, 2–10 (1973)

Gleiss N., E. Zwicker: Loudness Function in the Presence of Masking Noise. J. Acoust. Soc. Am. **36**, 393–394 (1964)

Goossens S., G. Bonin, R. Stumpner: Loudness perception with headphone presentation compared to loudspeaker presentation in the diffuse field. In *International Conference on Acoustics NAG/DAGA 2009*, 382–383 (Dt. Gesell. für Akustik e. V., Berlin, 2009)

Granier E., M. Kleiner, B.-I. Dalenbäck, P. Svensson: Experimental Auralization of Car Audio Installations. J. Audio Eng. Soc. **44**, 835–849 (1996)

Grantham D. W.: Detection and discrimination of simulated motion of auditory targets in the horizontal plane. J. Acoust. Soc. Am. **79**, 1939–1949 (1986)

Grantham D. W., B. W. Y. Hornsby, E. A. Erpenbeck: Auditory spatial resolution in horizontal, vertical, and diagonal planes. J. Acoust. Soc. Am. **114**, 1009–1022 (2003)

Greenfield R., M. O. Hawksford: Efficient Filter Design for Loudspeaker Equalization. J. Audio Eng. Soc. **39**, 739–751 (1991)

Greenwood D. D.: Critical Bandwidth and the Frequency Coordinates of the Basilar Membrane. J. Acoust. Soc. Am. **33**, 1344–1356 (1961)

Greenwood D. D.: A cochlear frequency-position function for several species – 29 years later. J. Acoust. Soc. Am. **87**, 2592–2605 (1990)

Griesinger D.: Binaural Techniques for Music Reproduction. In *8th International AES Conference* (1990) (Paper Number 8-026)

Griesinger D.: Frequency response adaptation in binaural hearing. In *126th AES Convention* (2009) (Convention Paper 7768)

Groh A. R.: High-Fidelity Sound System Equalization by Analysis of Standing Waves. J. Audio Eng. Soc. **22**, 795–799 (1974)

Gröhn M., T. Lokki, T. Takala: Localizing Sound Sources in a CAVE-Like Virtual Environment with Loudspeaker Array Reproduction. Presence **16**, 157–171 (2007)

Groth S., S. Merchel: Adaptive Adjustment of the "Sweet Spot" to the Listener's Position in a Stereophonic Play Back System – Part 2. In *International Conference on Acoustics NAG/DAGA 2009*, 1111–1114 (Dt. Gesell. für Akustik e. V., Berlin, 2009)

Habigt T., M. Durković, M. Rothbucher, K. Diepold: Enhancing 3D Audio using Blind Bandwidth Extension. In *129th AES Convention* (2010) (Convention Paper 8277)

Hacihabiboglu H., B. Gunel, F. Murtagh: Wavelet-Based Spectral Smoothing for Head-Related Transfer Function Filter Design. In *22nd International AES Conference* (2002) (Paper Number 246)

Haferkorn F., W. Schmid: System zur Erzeugung von vorgebbaren Hörereignisorten (System for the generation of predetermined auditory event positions). In *Fortschritte der Akustik, DAGA '96*, 366–367 (Dt. Gesell. für Akustik e. V., Oldenburg, 1996)

Hall J. L.: Minimum Detectable Change in Interaural Time or Intensity Difference for Brief Impulsive Stimuli. J. Acoust. Soc. Am. **36**, 2411–2413 (1964)

Hamasaki K., K. Hiyama, R. Okumura: The 22.2 Multichannel Sound System and Its Application. In *118th AES Convention* (2005) (Convention Paper 6406)

Hammershøi D., H. Møller: Sound transmission from the free field to the eardrum. Acustica – Acta Acustica **82**, S 87 (1996a)

Hammershøi D., H. Møller: Sound transmission to and within the human ear canal. J. Acoust. Soc. Am. **100**, 408–427 (1996b)

Hammershøi D., H. Møller: Methods for Binaural Recording and Reproduction. Acta Acustica united with Acustica **88**, 303–311 (2002)

Hammershøi D., J. Sandvad: Application of Binaural Synthesis for Auditory Virtual Environments. In *International Symposium on Simulation, Visualization and Auralization for Acoustic Research and Education*, 373–378 (1997)

Hartmann W. M., Z. A. Constan: Interaural level differences and the level-meter model. J. Acoust. Soc. Am. **112**, 1037–1045 (2002)

Hartmann W. M., B. Rakerd: On the minimum audible angle – A decision theory approach. J. Acoust. Soc. Am. **85**, 2031–2041 (1989)

Hatziantoniou P. D., J. N. Mourjopoulos: Addendum to "Generalized Fractional-Octave Smoothing of Audio and Acoustic Responses". J. Audio Eng. Soc. **48**, 940 (2000a)

Hatziantoniou P. D., J. N. Mourjopoulos: Generalized Fractional-Octave Smoothing of Audio and Acoustic Responses. J. Audio Eng. Soc. **48**, 259–280 (2000b)

Hawksford M. O. J.: Digital Signal Processing Tools for Loudspeaker Evaluation and Discrete-Time Crossover Design. J. Audio Eng. Soc. **45**, 37–62 (1997a)

Hawksford M. O. J.: High-definition digital audio in 3-dimensional sound reproduction. In *103rd AES Convention* (1997b) (Preprint 4560)

Hellbrück J., W. Ellermeier: *Hören – Physiologie, Psychologie und Pathologie (Hearing – physiology, psychology and pathology)*. 2nd Edition (Hogrefe, Göttingen, 2004)

Hellstrom P.-A., A. Axelsson: Miniature microphone probe tube measurements in the external auditory canal. J. Acoust. Soc. Am. **93**, 907–918 (1993)

Herre J.: From Joint Stereo to Spatial Audio Coding – Recent Progress and Standardization. In *7th International Conference on Digital Audio Effects (DAFX '04)* (2004)

Herre J., K. Brandenburg, D. Lederer: Intensity Stereo Coding. In *96th AES Convention* (1994) (Preprint 3799)

Hershkowitz R. M., N. I. Durlach: Interaural Time and Amplitude jnds for a 500-Hz Tone. J. Acoust. Soc. Am. **46**, 1464–1467 (1969)

Hertz B. F.: 100 Years with Stereo – The Beginning. In *68th AES Convention* (1981) (Preprint 1724)

Hesse A.: Comparison of Several Psychophysical Procedures with Respect to Threshold Estimates, Reproducibility and Efficiency. Acustica **59**, 263–273 (1986)

Hiekkanen T., A. Mäkivirta, M. Karjalainen: Virtualized Listening Tests for Loudspeakers. J. Audio Eng. Soc. **57**, 237–251 (2009)

Hirahara T.: Physical characteristics of headphones used in psychophysical experiments. Acoust. Sci. & Tech. **25**, 276–285 (2004)

Hoffmann P. F., H. Møller: Audibility of Spectral Switching in Head-Related Transfer Functions. In *119th AES Convention* (2005a) (Convention Paper 6537)

Hoffmann P. F., H. Møller: Audibility of Time Switching in Dynamic Binaural Synthesis. In *118th AES Convention* (2005b) (Convention Paper 6326)

Hoffmann P. F., H. Møller: Audibility of Spectral Differences in Head-Related Transfer Functions. In *120th AES Convention* (2006a) (Convention Paper 6652)

Hoffmann P. F., H. Møller: Audibility of time differences in adjacent head-related transfer functions. In *121st AES Convention* (2006b) (Convention Paper 6914)

Hoffmann P. F., H. Møller: Some Observations on Sensitivity to HRTF Magnitude. J. Audio Eng. Soc. **56**, 972–982 (2008)

Hojan E., H. Fastl: Intelligibility of Polish and German speech for the Polish audience in the presence of noise. Archives of Acoustics **21**, 123–130 (1996)

Hokari H., Y. Furumi, S. Shimada: A Study on Loudspeaker Arrangement in Multi-Channel Transaural System for Sound Image Localization. In *19th International AES Conference* (2001) (Paper Number 1922)

Horbach U.: New Techniques for the Production of Multichannel Sound. In *103rd AES Convention* (1997) (Preprint 4624)

Horbach U., M. M. Boone: Future Transmission and Rendering Formats for Multichannel Sound. In *16th International AES Conference* (1999) (Paper Number 16-036)

Horbach U., A. Karamustafaoglu, R. Pellegrini, P. Mackensen, G. Theile: Design and Applications of a Data-based Auralization System for Surround Sound. In *106th AES Convention* (1999) (Preprint 4976)

Horn T.: Image Processing of Speech with Auditory Magnitude Spectrograms. Acustica – Acta Acustica **84**, 175–177 (1998)

Houtgast T., S. Aoki: Stimulus-onset dominance in the perception of binaural information. Hearing Research **72**, 29–36 (1994)

Huang Y., J. Benesty, J. Chen: On Crosstalk Cancellation and Equalization With Multiple Loudspeakers for 3-D Sound Reproduction. IEEE Signal Processing Letters **14**, 649–652 (2007)

Hubel D. H., C. O. Henson, A. Rupert, R. Galambos: “Attention” Units in the Auditory Cortex. Science **129**, 1279–1280 (1959)

Hudde H., S. Schmidt: Sound fields in generally shaped curved ear canals. J. Acoust. Soc. Am. **125**, 3146–3157 (2009)

Huopaniemi J.: Future of Personal Audio – Smart Applications and Immersive Communication. In *30th International AES Conference* (2007) (Paper Number 34)

Hwang S., Y. Park: HRIR Customization in the Median Plane via Principal Components Analysis. In *31st International AES Conference* (2007) (Paper Number 9)

Hwang S., Y. Park, Y. Park: Customization of Spatially Continuous Head-Related Impulse Responses in the Median Plane. Acta Acustica united with Acustica **96**, 351–363 (2010)

Inanaga K., Y. Yamada, H. Koizumi: Headphone System with Out-of-Head Localisation Applying Dynamic HRTF (Head Related Transfer Function). In *98th AES Convention* (1995) (Preprint 4011)

Irino T., R. D. Patterson: A compressive gammachirp auditory filter for both physiological and psychophysical data. J. Acoust. Soc. Am. **109**, 2008–2022 (2001)

Irisawa H., S. Shimada, H. Hokari, S. Hosaya: Study of Fast Method to Calculate Inverse Filters. J. Audio Eng. Soc. **46**, 611–620 (1998)

Isdale J.: *What is Virtual Reality? A Homebrew Introduction.* `ftp://sunsite.unc.edu/pub/academic/computer-science/virtual-reality/papers/whatisvr.txt`. Version: March 1993, Visited: May 20, 2012

ITU-R BS.775-2: *Multichannel stereophonic sound system with and without accompanying picture.* (International Telecommunications Union, Geneva, July 2006)

Iwai T.: *Writing an ALSA Driver.* `http://ftp.task.gda.pl/mirror/ftp.kernel.org/pub/linux/kernel/people/tiwai/docs/writing-an-alsa-driver.pdf`. Version: February 2009, Visited: May 20, 2012

Iwaya Y.: Individualization of head-related transfer functions with tournament-style listening test: Listening with other's ears. Acoust. Sci. & Tech. **27**, 340–343 (2006)

Iwaya Y., Y. Suzuki, D. Kimura: Effects of head movement on front-back error in sound localization. Acoust. Sci. & Tech. **24**, 322–324 (2003)

Iwaya Y., Y. Suzuki, S. Takane: Effects of Listener's Head Movement on the Accuracy of Sound Localization in Virtual Environment. In *18th International Congress on Acoustics (ICA)*, II 997–1000 (2004)

Jeffress L. A., R. W. Taylor: Lateralization vs Localization. J. Acoust. Soc. Am. **33**, 482–483 (1961)

Jensen R. E., T. S. Welti: The Importance of Reflections in a Binaural Room Impulse Response. In *114th AES Convention* (2003) (Convention Paper 5839)

Jepsen M. L., S. D. Ewert, T. Dau: A computational model of human auditory signal processing and perception. J. Acoust. Soc. Am. **124**, 422–438 (2008)

Jin C., M. Schenkel, S. Carlile: Neural system identification model of human sound localization. J. Acoust. Soc. Am. **108**, 1215–1235 (2000)

Johnson D. H.: Origins of the Equivalent Circuit Concept: The Voltage-Source Equivalent. Proc. of the IEEE **91**, 636–640 (2003)

Jones M., S. J. Elliott, T. Takeuchi, J. Beer: Virtual Audio Reproduced in a Headrest. In *19th International Congress on Acoustics (ICA)* (2007)

Jot J.-M., V. Larcher, O. Warusfel: Digital Signal Processing in the Context of Binaural and Transaural Stereophony. In *98th AES Convention* (1995) (Preprint 3980)

Jürgens F., S. Werner: Kurven gleicher Lautheit bei binauraler Wiedergabe (Equal loudness contours for binaural presentation). In *Fortschritte der Akustik, DAGA 2012*, 217–218 (Dt. Gesell. für Akustik e. V., Berlin, 2012)

Jungmann J. O., R. Mazur, M. Kallinger, A. Mertins: Robust Combined Crosstalk Cancellation and Listening-Room Compensation. In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 9–12 (2011)

Kahana Y., P. A. Nelson, O. Kirkeby, H. Hamada: Objective and Subjective Assessment of Systems for the Production of Virtual Acoustic Images for Multiple Listeners. In *103rd AES Convention* (1997) (Preprint 4573)

Kahana Y., P. A. Nelson, O. Kirkeby, H. Hamada: A multiple microphone recording technique for the generation of virtual acoustic images. J. Acoust. Soc. Am. **105**, 1503–1516 (1999)

Kammeyer K. D.: *Nachrichtenübertragung (Information transmission).* (B. G. Teubner, Stuttgart, 1992)

Kammeyer K. D., K. Kroschel: *Digitale Signalverarbeitung (Digital signal processing).* 6th Edition (B. G. Teubner, Wiesbaden, 2006)

Katz B. F. G.: Acoustic absorption measurement of human hair and skin within the audible frequency range. J. Acoust. Soc. Am. **108**, 2238–2242 (2000)

Kayser H., S. D. Ewert, J. Anemüller, T. Rohdenburg, V. Hohmann, B. Kollmeier: Database of Multichannel In-Ear and Behind-the-Ear Head-Related and Binaural Room Impulse Responses. EURASIP Journal on Advances in Signal Processing **2009**, Article ID 298605 (2009)

Keele D. B.: A Loudspeaker Horn that Covers a Flat Rectangular Area from an Oblique Angle. In *74th AES Convention* (1983) (Preprint 2052)

Keidser G., R. Katsch, H. Dillon, F. Grant: Relative loudness perception of low and high frequency sounds in the open and occluded ear. J. Acoust. Soc. Am. **107**, 3351–3357 (2000)

Keyrouz F., K. Diepold: A New HRTF Interpolation Approach for Fast Synthesis of Dynamic Environmantal Interaction. J. Audio Eng. Soc. **56**, 28–35 (2008)

Killion M. C.: Revised estimate of minimum audible pressure: Where is the ”missing 6 dB”? J. Acoust. Soc. Am. **63**, 1501–1508 (1978)

Kim S.-M., W. Choi: On the externalization of virtual sound images in headphone reproduction: A Wiener filter approach. J. Acoust. Soc. Am. **117**, 3657–3665 (2005)

Kim S.-M., S. Wang: A Wiener filter approach to the binaural reproduction of stereo sound. J. Acoust. Soc. Am. **114**, 3179–3188 (2003)

Kim S., D. Kong, S. Jang: Adaptive Virtual Surround Sound Rendering System for an Arbitrary Listening Position. J. Audio Eng. Soc. **56**, 243–254 (2008)

Kirkeby O., P. A. Nelson: Digital Filter Design for Virtual Source Imaging Systems. In *104th AES Convention* (1998) (Preprint 4688)

Kirkeby O., P. A. Nelson: Digital Filter Design for Inversion Problems in Sound Reproduction. J. Audio Eng. Soc. **47**, 583–595 (1999)

Kirkeby O., P. Rubak, A. Farina: Analysis of Ill-Conditioning of Multi-Channel Deconvolution Problems. In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 155–158 (1999)

Kirkeby O., P. A. Nelson, H. Hamada, F. Orduna-Bustamante: Fast Deconvolution of Multichannel Systems Using Regularization. IEEE Transactions on Speech and Audio Processing **6**, 189–194 (1998)

Kistler D. J., F. L. Wightman: A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction. J. Acoust. Soc. Am. **91**, 1637–1647 (1992)

Klehs B., T. Sporer: Wave Field Synthesis in the Real World: Part 1 – In the Living Room. In *114th AES Convention* (2003) (Convention Paper 5727)

Klipsch P. W.: Intensity vs Time Differences in Stereo. J. Audio Eng. Soc. **8**, 139 (1960)

König F. M.: 4-Canal Headphone for in Front Localization and HDTV- or Dolby-Surround Use. In *96th AES Convention* (1994) (Preprint 3826)

König F. M.: New Measurements and Psycho-Acoustic Investigations on a Headphone for TAX / HDTV / Dolby-Surround Reproduction of Sound. In *98th AES Convention* (1995) (Preprint 4010)

Kopčo N., B. G. Shinn-Cunningham: Spatial unmasking of nearby pure-tone targets in a simulated anechoic environment. J. Acoust. Soc. Am. **114**, 2856–2870 (2003)

Kopčo N., B. G. Shinn-Cunningham: Influences of modulation and spatial separation on detection of a masked broadband target. J. Acoust. Soc. Am. **124**, 2236–2250 (2008)

Köring J., A. Schmitz: Simplifying Cancellation of Cross-Talk for Playback of Head-Related Recordings in a Two-Speaker System. Acustica **79**, 221–232 (1993)

Krumbholz K., R. D. Patterson, A. Nobbe, H. Fastl: Microsecond temporal resolution in monaural hearing without spectral cues? J. Acoust. Soc. Am. **113**, 2790–2800 (2003)

Kulkarni A., H. S. Colburn: Role of spectral detail in sound-source localization. Nature **396**, 747–749 (1998)

Kulkarni A., H. S. Colburn: Variability in the characterization of the headphone transfer-function. J. Acoust. Soc. Am. **107**, 1071–1074 (2000)

Kulkarni A., S. K. Isabelle, H. S. Colburn: Sensitivity of human subjects to head-related transfer-function phase spectra. J. Acoust. Soc. Am. **105**, 2821–2840 (1999)

Kuttruff H., M. J. Jusofie: Nachhallmessungen nach dem Verfahren der integrierten Impulsantwort (Measurements of reverberation based on the method of the integrated impulse response). Acustica **19**, 56–58 (1967/68)

Kuttruff H., M. J. Jusofie: Messungen des Nachhallverlaufes in mehreren Räumen, ausgeführt nach dem Verfahren der integrierten Impulsantwort (Measurements of the reverberation decay in several rooms, carried out based on the method of the integrated impulse response). Acustica **21**, 1–9 (1969)

Langendijk E. H. A., A. W. Bronkhorst: Fidelity of three-dimensional-sound reproduction using a virtual auditory display. J. Acoust. Soc. Am. **107**, 528–537 (2000)

Larsson P., D. Västfjäll, M. Kleiner: Effects of auditory information consistency and room acoustic cues on presence in virtual environments. Acoust. Sci. & Tech. **29**, 191–194 (2008)

Lavie N., A. Hirst, J. W. de Fockert, E. Viding: Load Theory of Selective Attention and Cognitive Control. J. Exp. Psychol.: General **133**, 339–354 (2004)

Laws P.: Experimentelle Untersuchungen zur Im-Kopf-Lokalisiertheit (IKL) von Hörereignissen (Experimental investigations on the inside-the-head locatedness of auditory events). In *Fortschritte der Akustik, DAGA '72*, 329–332 (VDE, Berlin, 1972a)

Laws P.: *Untersuchungen zum Entfernungshören und zum Problem der Im-Kopf-Lokalisiertheit von Hörereignissen (Studies on distance hearing and on the problem of the inside-the-head localization of auditory events)*, PhD thesis, Rheinisch-Westfälisch Technische Hochschule Aachen (1972b)

Laws P., J. Blauert: Ein Beitrag zur Hörbarkeit von Laufzeitverzerrungen (A contribution on the audibility of group delay distortions). In *Fortschritte der Akustik, DAGA '73*, 447–450 (VDI, Düsseldorf, 1973)

Leakey D. M.: Some Measurements on the Effects of Interchannel Intensity and Time Differences in Two Channel Sound Systems. J. Acoust. Soc. Am. **31**, 977–986 (1959)

Lee H.: The Relationship between Interchannel Time and Level Differences in Vertical Sound Localisation and Masking. In *131st AES Convention* (2011) (Convention Paper 8556)

Lee K., C. Son, D. Kim: Immersive Virtual Sound Beyond 5.1 Channel Audio. In *128th AES Convention* (2010) (Convention Paper 8117)

Leitner S., A. Sontacchi, R. Höldrich: Multichannel Sound Reproduction System for Binaural Signals – The Ambisonic Approach. In 3rd *International Conference on Digital Audio Effects (DAFX '00)* (2000)

Lentz T.: Dynamic Crosstalk Cancellation for Binaural Synthesis in Virtual Reality Environments. J. Audio Eng. Soc. **54**, 283–294 (2006)

Lentz T., G. Behler: Dynamic Cross-Talk Cancellation for Binaural Synthesis in Virtual Reality Environments. In *117th AES Convention* (2004) (Convention Paper 6315)

Lentz T., O. Schmitz: Realisierung eines Echtzeit-Systems zur Nachführung der Übersprechkompensation für einen bewegten Zuhörer (Realization of a realtime system for the adaptation of crosstalk cancellation to a moving listener). In *Fortschritte der Akustik, DAGA '02*, 730–731 (Dt. Gesell. für Akustik e. V., Oldenburg, 2002)

Levitt H.: Transformed Up-Down Methods in Psychoacoustics. J. Acoust. Soc. Am. **49**, 467–477 (1971)

Lindau A., H.-J. Maempel, S. Weinzierl: Minimum BRIR grid resolution for dynamic binaural synthesis. In *Acoustics '08*, 3851–3856 (2008)

Lindner F., F. Völk, H. Fastl: Simulation und psychoakustische Bewertung von Übertragungsfehlern bei der Wellenfeldsynthese (Simulation and psychoacoustic assessment of transmission errors of wave field synthesis). In *Fortschritte der Akustik, DAGA 2011*, 663–664 (Dt. Gesell. für Akustik e. V., Berlin, 2011)

Loizou P. C.: Mimicking the Human Ear – An Overview of Signal-Processing Strategies for Converting Sound into Electrical Signals in Cochlear Implants. IEEE Signal Processing Magazine **15**, 101–130 (1998)

Lokki T.: Subjective comparison of four concert halls based on binaural impulse responses. Acoust. Sci. & Tech. **26**, 200–203 (2005)

Loomis J. M., R. G. Golledge, R. L. Klatzky: Navigation System for the Blind: Auditory Display Modes and Guidance. Presence **7**, 193–203 (1998)

Mackensen P., K. Reichenauer, G. Theile: Einfluß der spontanen Kopfdrehungen auf die Lokalisation beim binauralen Hören (Influence of spontaneous head rotations on the localization in binaural hearing). In *20. Tonmeistertagung, VDT International Convention*, 218–228 (Verband deutscher Tonmeister, VDT, 1998)

Mackensen P., U. Felderhoff, G. Theile, U. Horbach, R. Pellegrini: Binaural Room Scanning – A new Tool for Acoustic and Psychoacoustic Research. In *137th Meeting of the Acoustical Society of America* (Acoustical Society of America, Melville, NY, 1999)

Macpherson E. A., J. C. Middlebrooks: Listener weighting of cues for lateral angle: The duplex theory of sound localization revisited. J. Acoust. Soc. Am. **111**, 2219–2236 (2002)

Martens W. L.: Perceptual evaluation of filters controlling source direction: Customized and generalized HRTFs for binaural synthesis. Acoust. Sci. & Tech. **24**, 220–232 (2003)

Martens W. L., A. Guru, D. Lee: Effects of individualised headphone response equalization on front/back hemifield discrimination for virtual sources displayed on the horizontal plane. In *20th International Congress on Acoustics (ICA)* (2010)

Martignon P., A. Azzali, D. Cabrera, A. Capra, A. Farina: Reproduction of auditorium spatial impression with binaural and stereophonic sound systems. In *118th AES Convention* (2005) (Convention Paper 6485)

McAnally K. I., R. I. Martin: Variability in the Headphone-to-Ear-Canal Transfer Function. J. Audio Eng. Soc. **50**, 263–266 (2002)

Meares D. J., G. Theile: Matrixed Surround Sound in an MPEG Digital World. J. Audio Eng. Soc. **46**, 331–335 (1998)

Meesawat K., D. Hammershøi: The time when the reverberation tail in a binaural room impulse response begins. In *115th AES Convention* (2003) (Convention Paper 5859)

Mehrgardt S.: Die Übertragungsfunktion des menschlichen Außenohres – Richtungsabhängigkeit und genauere Bestimmung durch komplexe Strukturmittelung (The human outer ear transfer function – Dependence on direction and more accurate determination by complex structure averaging). In *Fortschritte der Akustik, DAGA '75*, 357–360 (Physik, Weinheim, 1975)

Mehrgardt S., V. Mellert: Transformation characteristics of the external human ear. J. Acoust. Soc. Am. **61**, 1567–1576 (1977)

Melchior F., J.-O. Fischer, D. de Vries: Audiovisual Perception using Wave Field Synthesis in Combination with Augmented Reality Systems: Horizontal Positioning. In *28th International AES Conference* (2006) (Paper Number 3-2)

Menzel D.: *Zum Einfluss von Farben auf das Lautheitsurteil (On the influence of colors on the loudness judgement)*, PhD thesis, Technische Universität München (Dr. Hut, München, 2011)

Menzel D., H. Fastl, T. Brandt, T. Stephan: An active free-field equalizer for headphones used in functional magnetic resonance imaging. Acoust. Sci. & Tech. **32**, 251–254 (2011a)

Menzel D., H. Fastl, R. Graf, J. Hellbrück: Influence of vehicle color on loudness judgments. J. Acoust. Soc. Am. **123**, 2477–2479 (2008)

Menzel D., H. Wittek, H. Fastl, G. Theile: Binaurale Raumsynthese mittels Wellenfeldsynthese – Realisierung und Evaluierung (Binaural room synthesis by wave field synthesis – realization and evaluation). In *Fortschritte der Akustik, DAGA '06*, 255–256 (Dt. Gesell. für Akustik e. V., Berlin, 2006)

Menzel D., H. Wittek, G. Theile, H. Fastl: The Binaural Sky: A Virtual Headphone for Binaural Room Synthesis. In *1st International VDT Symposium* (Verband deutscher Tonmeister, VDT, 2005)

Menzel D., A. Gottschalk, N. Haufe, F. Völk, H. Fastl: Zum Einfluss der optischen Darbietungsweise auf audio-visuelle Interaktionen beim Lautheitsurteil (On the influence of the optical presentation method on audio-visual interactions regarding the loudness judgment). In *Fortschritte der Akustik, DAGA 2011*, 587–588 (Dt. Gesell. für Akustik e. V., Berlin, 2011b)

Merchel S., S. Groth: Adaptive Adjustment of the “Sweet Spot” to the Listener's Position in a Stereophonic Play Back System – Part 1. In *International Conference on Acoustics NAG/DAGA 2009*, 1093–1095 (Dt. Gesell. für Akustik e. V., Berlin, 2009a)

Merchel S., S. Groth: Analysis and Implementation of a Stereophonic Play Back System for Adjusting the “Sweet Spot” to the Listener's Position. In *126th AES Convention* (2009b) (Convention Paper 7726)

Merchel S., S. Groth: Adaptive Adjustment of the “Sweet Spot” for Head Rotation. In *20th International Congress on Acoustics (ICA)* (2010a)

Merchel S., S. Groth: Automatische Anpassung des stereophonen Sweetspots bei Kopfdrehung (Automatic adaption of the stereophonic sweet spot on head rotation). In *Fortschritte der Akustik, DAGA 2010*, 635–636 (Dt. Gesell. für Akustik e. V., Berlin, 2010b)

Microsoft Corp.: *HD Audio Guidelines for Windows.* September 2, 2009 Edition. Redmond, WA, USA (September 2009)

Middlebrooks J. C.: Individual differences in external-ear transfer functions reduced by scaling in frequency. J. Acoust. Soc. Am. **106**, 1480–1492 (1999a)

Middlebrooks J. C.: Virtual localization improved by scaling nonindividualized external-ear transfer functions in frequency. J. Acoust. Soc. Am. **106**, 1493–1510 (1999b)

Middlebrooks J. C., D. M. Green: Sound Localization by Human Listeners. Annu. Rev. Psychol. **42**, 135–159 (1991)

Middlebrooks J. C., E. A. Macpherson, Z. A. Onsan: Psychophysical customization of directional transfer functions for virtual sound localization. J. Acoust. Soc. Am. **108**, 3088–3091 (2000)

Middlebrooks J. C., J. C. Makous, D. M. Green: Directional sensitivity of sound-pressure levels in the human ear canal. J. Acoust. Soc. Am. **86**, 89–108 (1989)

Mills A. W.: On the Minimum Audible Angle. J. Acoust. Soc. Am. **30**, 237–246 (1958)

Mills A. W.: *Chapter* Auditory Localization. In *Foundations of Modern Auditory Theory, Vol. II*, ed. by J. V. Tobias, 303–348 (Academic Press, New York, 1972)

Minnaar P., J. Plogsties, F. Christensen: Directional Resolution of Head-Related Transfer Functions Required in Binaural Synthesis. J. Audio Eng. Soc. **53**, 919–929 (2005)

Minnaar P., S. K. Olesen, F. Christensen, H. Møller: Localization with Binaural Recordings from Artificial and Human Heads. J. Audio Eng. Soc. **49**, 323–336 (2001)

Minnaar P., S. K. Olesen, F. Christensen, H. Møller: Sound localisation with binaural recordings made with artificial heads. In *18th International Congress on Acoustics (ICA)*, V 3651–3654 (2004)

Moldrzyk C., A. Goertz, M. Makarski, S. Feistel, W. Ahnert, S. Weinzierl: Wellenfeldsynthese für einen großen Hörsaal (Wave field synthesis for a big lecture hall). In *Fortschritte der Akustik, DAGA 2007*, 683–684 (Dt. Gesell. für Akustik e. V., Berlin, 2007)

Møller H.: Cancellation of crosstalk in artificial head recordings, reproduced through loudspeakers. In *84th AES Convention* (1988) (Preprint 2610)

Møller H.: Reproduction of Artificial-Head Recordings through Loudspeakers. J. Audio Eng. Soc. **37**, 30–33 (1989)

Møller H.: Fundamentals of Binaural Technology. Applied Acoustics **36**, 171–218 (1992)

Møller H., D. Hammershøi, C. B. Jensen, M. F. Sørensen: Transfer Characteristics of Headphones Measured on Human Ears. J. Audio Eng. Soc. **43**, 203–217 (1995a)

Møller H., D. Hammershøi, C. B. Jensen, M. F. Sørensen: Evaluation of Artificial Heads in Listening Tests. J. Audio Eng. Soc. **47**, 83–100 (1999)

Møller H., C. B. Jensen, D. Hammershøi, M. F. Sørensen: Selection of a typical human subject for binaural recording. Acustica – Acta Acustica **82**, S 215 (1996a)

Møller H., C. B. Jensen, D. Hammershøi, M. F. Sørensen: Using a Typical Human Subject for Binaural Recording. In *100th AES Convention* (1996b) (Preprint 4157)

Møller H., C. B. Jensen, D. Hammershøi, M. F. Sørensen: Evaluation of Artificial Heads in Listening Tests. In *102nd AES Convention* (1997) (Preprint 4404)

Møller H., M. F. Sørensen, C. B. Jensen, D. Hammershøi: Binaural Technique: Do We Need Individual Recordings? J. Audio Eng. Soc. **44**, 451–469 (1996c)

Møller H., M. F. Sørensen, D. Hammershøi, C. B. Jensen: Head-Related Transfer Functions of Human Subjects. J. Audio Eng. Soc. **43**, 300–321 (1995b)

Mouchtaris A., P. Reveliotis, C. Kyriakakis: Non-minimum phase inverse filter methods for immersive audio rendering. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 3077–3080 (1999)

Mouchtaris A., P. Reveliotis, C. Kyriakakis: Inverse Filter Design for Immersive Audio Rendering Over Loudspeakers. IEEE Transactions on Multimedia **2**, 77–87 (2000)

Mourjopoulos J.: Digital equalization methods for audio systems. In *84th AES Convention* (1988) (Preprint 2598)

Mourjopoulos J., P. M. Clarkson, J. K. Hammond: A Comparative Study of Least-Squares and Homomorphic Techniques for the Inversion of Mixed Phase Signals. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1858–1861 (1982)

Müller S., P. Massarani: Transfer-Function Measurement with Sweeps. J. Audio Eng. Soc. **49**, 443–471 (2001)

Mummert M.: *Sprachcodierung durch Konturierung eines gehörangepaßten Spektrogramms und ihre Anwendung zur Datenreduktion (Speech coding by contouring an ear-adapted spectrogram and its application to data reduction)*, PhD thesis, Technische Universität München (1997)

Munson W. A., F. M. Wiener: In Search of the Missing 6 Db. J. Acoust. Soc. Am. **24**, 498–501 (1952)

Nathanail C., C. Lavandier, J.-D. Polack: Influence of the visual information on auditory perception. Consequences on the objective characterization of room acoustics. Acustica – Acta Acustica **82**, S 216 (1996)

Neely S. T., J. B. Allen: Invertibility of a room impulse response. J. Acoust. Soc. Am. **66**, 165–169 (1979)

Neely S. T., M. P. Gorga: Comparison between intensity and pressure as measures of sound level in the ear canal. J. Acoust. Soc. Am. **104**, 2925–2934 (1998)

Nicol R., M. Emerit: Reproducing 3D-Sound for Videoconferencing: a Comparison between Holophony and Ambisonic. In 1*st* *International Conference on Digital Audio Effects (DAFX '98)* (1998)

Norcross S. G., M. Bouchard, G. A. Soulodre: Adaptive Strategies for Inverse Filtering. In *119th AES Convention* (2005) (Convention Paper 6563)

Norcross S. G., M. Bouchard, G. A. Soulodre: Inverse Filtering Design Using a Minimal-Phase Target Function from Regularization. In *121st AES Convention* (2006) (Convention Paper 6929)

Norcross S. G., G. A. Soulodre, M. C. Lavoie: Evaluation of Inverse Filtering Techniques for Room/Speaker Equalization. In *113th AES Convention* (2002) (Convention Paper 5662)

Norcross S. G., G. A. Soulodre, M. C. Lavoie: Further Investigations of Inverse Filtering. In *115th AES Convention* (2003a) (Convention Paper 5923)

Norcross S. G., G. A. Soulodre, M. C. Lavoie: Subjective Effects of Regularization on Inverse Filtering. In *114th AES Convention* (2003b) (Convention Paper 5848)

Norcross S. G., G. A. Soulodre, M. C. Lavoie: Distortion Audibility in Inverse Filtering. In *117th AES Convention* (2004a) (Convention Paper 6311)

Norcross S. G., G. A. Soulodre, M. C. Lavoie: Subjective Investigations of Inverse Filtering. J. Audio Eng. Soc. **52**, 1003–1027 (2004b)

Nyquist H.: Certain Factors Affecting Telegraph Speed. Bell System Technical Journal **3**, 324–346 (1924)

O'Dwyer M. F., G. Potard, I. Burnett: A 16-Speaker 3D Audio-Visual Display Interface and Control System. In *10th International Conference on Auditory Display (ICAD)* (2004)

Olson H. F.: Microphones for Recording. J. Audio Eng. Soc. **25**, 676–684 (1977)

Oppenheim A. V., R. W. Schafer, J. R. Buck: *Discrete-Time Signal Processing.* 2nd Edition (Prentice-Hall, New Jersey, 1999)

Oppenheim A. V., A. S. Willsky, H. Nawab: *Signals and Systems.* 2nd Edition (Prentice-Hall, New Jersey, 1998)

Otani M., T. Hirahara, S. Ise: Numerical study on source-distance dependency of head-related transfer functions. J. Acoust. Soc. Am. **125**, 3253–3261 (2009)

Otto N. C.: Listening Test Methods for Automotive Sound Quality. In *103rd AES Convention* (1997) (Preprint 4586)

Panzer J., L. Ferekidis: The use of continuous phase for interpolation, smoothing and forming mean values of complex frequency response curves. In *116th AES Convention* (2004) (Convention Paper 6005)

Patterson R. D., K. Robinson, J. Holdsworth, D. McKeown, C. Zhang, M. Allerhand: Complex sounds and auditory images. In 9*th* *International Symposium on Hearing*, ed. by Y. Cazals, L. Demany, K. Horner, 429–446 (Pergamon, Oxford, 1992)

Peisl W.: *Beschreibung aktiver nichtlinearer Effekte der peripheren Schallverarbeitung des Gehörs durch ein Rechnermodell (Describing active nonlinear effects of the peripheral sound-processing of the hearing system by a computer-model)*, PhD thesis, Technische Universität München (1990)

Pellegrini R. S., U. Horbach: Perceptual Encoding of Acoustic Environments. In *IEEE International Conference on Multimedia and Expo (ICME)*, 501–503 (2002)

Pellegrini R. S., C. Kuhn: Augmented Reality Using Wave Field Synthesis for Theatres and Opera Houses. In *Fortschritte der Akustik, DAGA '05*, 203–204 (Dt. Gesell. für Akustik e. V., Berlin, 2005)

Pellegrini R. S., C. Kuhn, M. Gebhardt: Headphones Technology for Surround Sound Monitoring – A Virtual 5.1 Listening Room. In *122nd AES Convention* (2007) (Convention Paper 7068)

Perrott D. R.: Auditory and Visual Localization: Two Modalities, One World. In *12th International AES Conference*, 221–231 (1993) (Paper Number 12-018)

Perrott D. R., A. D. Musicant: Minimum auditory movement angle: Binaural localization of moving sound sources. J. Acoust. Soc. Am. **62**, 1463–1466 (1977)

Perrott D. R., S. Pacheco: Minimum audible angle thresholds for broadband noise as a function of the delay between the onset of the lead and lag signals. J. Acoust. Soc. Am. **85**, 2669–2672 (1989)

Perrott D. R., K. Saberi: Minimum audible angle thresholds for sources varying in both elevation and azimuth. J. Acoust. Soc. Am. **87**, 1728–1731 (1990)

Perrott D. R., J. Tucker: Minimum audible movement angle as a function of signal frequency and the velocity of the source. J. Acoust. Soc. Am. **83**, 1522–1527 (1988)

Plack C. J., B. C. J. Moore: Temporal window shape as a function of frequency and level. J. Acoust. Soc. Am. **87**, 2178–2187 (1990)

Platte H.-J., K. Genuit: Ein Beitrag zum Verständnis der Summenlokalisation (A contribution to the understanding of summing localization). In *Fortschritte der Akustik, DAGA '80*, 595–598 (VDE, Berlin, 1980)

Plenge G.: On the differences between localization and lateralization. J. Acoust. Soc. Am. **56**, 944–951 (1974)

Plogsties J., S. K. Olesen, P. Minnaar, F. Christensen, H. Møller: Audibility of All-Pass Components in Head-Related Transfer Functions. In *108th AES Convention* (2000) (Preprint 5132)

Poitschke T., F. Laquai, G. Rigoll: Guiding a Driver's Visual Attention Using Graphical and Auditory Animations. In *Engin. Psychol. and Cog. Ergonomics, HCII*, 424–433 (Springer, Berlin, Heidelberg, 2009)

Polhemus: *3SPACE™ FASTRAK® User's Manual.* OPM00PI002 Rev. E Edition. Colchester, Vermont, USA (April 2005)

Pörschmann C.: 3-D Audio in mobilen Kommunikationsendgeräten (3-D audio in mobile communication terminals). In *Fortschritte der Akustik, DAGA '02*, 732–733 (Dt. Gesell. für Akustik e. V., Oldenburg, 2002)

Port E.: Ein elektrostatischer Lautsprecher zur Erzeugung eines ebenen Schallfeldes (An electrostatic loudspeaker for the generation of a plain sound field). Frequenz **18**, 9–13 (1964)

Potchinkov A.: Frequency-Domain Equalization of Audio Systems Using Digital Filters – Part 1: Basics of Filter Design. J. Audio Eng. Soc. **46**, 977–987 (1998a)

Potchinkov A.: Frequency-Domain Equalization of Audio Systems Using Digital Filters – Part 2: Examples of Equalization. J. Audio Eng. Soc. **46**, 1092–1108 (1998b)

Pralong D., S. Carlile: The role of individualized headphone calibration for the generation of high fidelity virtual auditory space. J. Acoust. Soc. Am. **100**, 3785–3793 (1996)

Preis D.: Phase Distortion and Phase Equalization in Audio Signal Processing – A Tutorial Review. J. Audio Eng. Soc. **30**, 774–794 (1982)

Pritchett J.: The Subjective Free-Field Calibration of an External Telephone Receiver by the Equal-Loudness Method. Acustica **4**, 544–546 (1954)

Pulkki V.: Virtual Sound Source Positioning Using Vector Base Amplitude Panning. J. Audio Eng. Soc. **45**, 456–466 (1997)

Pulkki V.: Localization of Amplitude-Panned Virtual Sources II: Two- and Three-Dimensional Panning. J. Audio Eng. Soc. **49**, 753–767 (2001)

Pulkki V., M. Karjalainen: Localization of Amplitude-Panned Virtual Sources I: Stereophonic Panning. J. Audio Eng. Soc. **49**, 739–752 (2001)

Qin M. K., A. J. Oxenham: Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers. J. Acoust. Soc. Am. **114**, 446–454 (2003)

Rader T., C. Schmiegelow, U. Baumann, H. Fastl: Oldenburger Satztest im „Multi-Source Noise Field“ mit unterschiedlichen Modulationscharakteristika (Oldenburger sentence test in the “Multi-Source Noise Field” with different modulation characteristics). In *Fortschritte der Akustik, DAGA 2008*, 663–664 (Dt. Gesell. für Akustik e. V., Berlin, 2008)

Rao H. I. K., V. J. Mathews, Y.-C. Park: A Minimax Approach for the Joint Design of Acoustic Crosstalk Cancellation Filters. IEEE Transactions on Audio, Speech, and Language Processing **15**, 2287–2298 (2007)

Rayleigh L.: On Pin-hole Photography. Philosophical Magazine **31**, 87–99 (1891)

Richmond S. A., J. G. Kopun, S. T. Neely, H. Tan, M. P. Gorga: Distribution of standing-wave errors in real-ear sound-level measurements. J. Acoust. Soc. Am. **129**, 3134–3140 (2011)

RME – Intelligent Audio Solutions: *Fireface 400 User's Guide.* 1.8 Edition. Audio AG, Am Pfanderling 60, 85778 Haimhausen (February 2011)

Robinson D. W.: The Relation between the Sone and Phon Scales of Loudness. Acustica **3**, 344–358 (1953)

Robinson D. W., R. S. Dadson: A re-determination of the equal-loudness relations for pure tones. Brit. J. of Appl. Physics **7**, 166–181 (1956)

Roman N., D. Wang, G. J. Brown: Speech segregation based on sound localization. J. Acoust. Soc. Am. **114**, 2236–2252 (2003)

Rose J., P. Nelson, B. Rafaely, T. Takeuchi: Sweet spot size of virtual acoustic imaging systems at asymmetric listener locations. J. Acoust. Soc. Am. **112**, 1992–2002 (2002)

Roys H. E.: The Coming of Stereo. J. Audio Eng. Soc. **25**, 824–827 (1977)

Rudmose W.: Free-Field Thresholds vs. Pressure Thresholds at Low Frequencies. J. Acoust. Soc. Am. **22**, 674 (1950)

Rudmose W.: The case of the missing 6 dB. J. Acoust. Soc. Am. **71**, 650–659 (1982)

Ryan C., D. Furlong: Effects of Headphone Placement on Headphone Equalisation for Binaural Reproduction. In *98th AES Convention* (1995) (Preprint 4009)

Saberi K., L. Dostal, T. Sadralodabai, D. R. Perrott: Minimum Audible Angles for Horizontal, Vertical, and Oblique Orientations: Lateral and Dorsal Planes. Acustica **75**, 57–61 (1991)

Sandvad J.: Dynamic Aspects of Auditory Virtual Environments. In *100th AES Convention* (1996) (Preprint 4226)

Savioja L.: *Modeling Techniques for Virtual Acoustics*, PhD thesis, Helsinki University of Technology (2000)

Schlang M., M. Mummert: Die Bedeutung der Fensterfunktion für die Fourier-t-Transformation als gehörgerechte Spektralanalyse (The relevance of the window function for the Fourier-t-transform as auditory adequate spectral analysis method). In *Fortschritte der Akustik, DAGA '90*, 1043–1046 (DPG, Bad Honnef, 1990)

Schmidhuber M., F. Völk, H. Fastl: Psychoakustische Experimente zum Einfluss des Ventriloquismuseffekts auf Richtungsunterschiedsschwellen (Minimum Audible Angles) in der Horizontalebene (Psychoacoustic experiments on the influence of the ventriloquism effect on minimum audible angles in the horizontal plane). In *Fortschritte der Akustik, DAGA 2011*, 577–578 (Dt. Gesell. für Akustik e. V., Berlin, 2011)

Schmidt S., H. Hudde: Accuracy of acoustic ear canal impedances: Finite element simulation of measurement methods using a coupling tube. J. Acoust. Soc. Am. **125**, 3819–3827 (2009)

Schneider B., S. Parker: Does stimulus context affect loudness or only loudness judgments? Perception & Psychophysics **48**, 409–418 (1990)

Schönstein D., B. F. G. Katz: HRTF selection for binaural synthesis from a database using morphological parameters. In *20th International Congress on Acoustics (ICA)* (2010)

Schorer E.: Critical modulation frequency based on detection of AM versus FM tones. J. Acoust. Soc. Am. **79**, 1054–1057 (1986)

Schorer E.: *Ein Funktionsschema zur Beschreibung eben wahrnehmbarer Frequenz- und Amplitudenänderungen (A functional scheme describing just noticeable frequency and amplitude variations)*, PhD thesis, Technische Universität München (1988)

Schorer E.: Vergleich eben erkennbarer Unterschiede und Variationen der Frequenz und Amplitude von Schallen (Comparison of just noticeable differences and variations of frequency and amplitude of sounds). Acustica **68**, 183–199 (1989)

Schroeder M. R.: New Method of Measuring Reverberation Time. J. Acoust. Soc. Am. **37**, 409–412 (1965)

Schroeder M. R., B. S. Atal, G. M. Sessler, J. E. West: Acoustical Measurements in Philharmonic Hall (New York). J. Acoust. Soc. Am. **40**, 434–440 (1966)

Schwartz A. H., B. G. Shinn-Cunningham: Dissociation of perceptual judgments of "what" and "where" in an ambiguous auditory scene. J. Acoust. Soc. Am. **128**, 3041–3051 (2010)

Scriven E. O.: Symposium on Wire Transmission of Symphonic Music and Its Reproduction in Auditory Perspective – Amplifiers. Bell System Technical Journal **13**, 278–284 (1934)

Seeber B. U.: Untersuchung der Lokalisation in reflexionsarmer Umgebung und bei virtueller akustischer Richtungsdarbietung mit einer Laser-Pointer-Methode (Investigation of localization in anechoic environment and with virtual acoustic direction presentation using a laser pointer method). In *Fortschritte der Akustik, DAGA '02*, 482–483 (Dt. Gesell. für Akustik e. V., Oldenburg, 2002a)

Seeber B. U.: Zum Ventriloquismus-Effekt in realer und virtueller Hörumgebung (On the ventriloquism effect in real and virtual listening environment). In *Fortschritte der Akustik, DAGA '02*, 480–481 (Dt. Gesell. für Akustik e. V., Oldenburg, 2002b)

Seeber B. U.: *Untersuchung der auditiven Lokalisation mit einer Lichtzeigermethode (Investigation of the auditory localization using a light pointer method)*, PhD thesis, Technische Universität München (2003)

Seeber B. U.: The Duplex-Theory of Localization Investigated under Natural Conditions. In *19th International Congress on Acoustics (ICA)* (2007)

Seeber B. U.: Weighting of binaural cues in the absence of a reflection. In *Fortschritte der Akustik, DAGA 2010*, 641–642 (Dt. Gesell. für Akustik e. V., Berlin, 2010)

Seeber B. U., H. Fastl: Effiziente Auswahl der individuell-optimalen aus fremden Außenohrübertragungsfunktionen (Efficient selection of the optimal individual from nonindividual head-related transfer functions). In *Fortschritte der Akustik, DAGA '01*, 484–485 (Dt. Gesell. für Akustik e. V., Oldenburg, 2001)

Seeber B. U., H. Fastl: Subjective Selection of Non-Individual Head-Related Transfer Functions. In *9th International Conference on Auditory Display (ICAD)* (2003)

Seeber B. U., E. Hafter: Perceptual equalization in near-speaker panning. In *Fortschritte der Akustik, DAGA 2007*, 375–376 (Dt. Gesell. für Akustik e. V., Berlin, 2007)

Seeber B. U., S. Kerber, E. R. Hafter: A system to simulate and reproduce audio-visual environments for spatial hearing research. Hearing Research **260**, 1–10 (2010)

Shannon C. E.: Communication in the Presence of Noise. Proc. Inst. of Radio Eng. **37**, 10–21 (1949)

Shaw E. A. G.: Earcanal Pressure Generated by Circumaural and Supraaural Earphones. J. Acoust. Soc. Am. **39**, 471–479 (1966)

Shimada S., S. Hayashi: Stereophonic Sound Image Localization System Using Inner-Earphones. Acustica **81**, 264–271 (1995)

Silzle A.: Auswahl und Tuning von HRTFs (Selection and Tuning of HRTFs). In *Fortschritte der Akustik, DAGA '02*, 734–735 (Dt. Gesell. für Akustik e. V., Oldenburg, 2002a)

Silzle A.: Selection and Tuning of HRTFs. In *112th AES Convention* (2002b) (Convention Paper 5595)

Silzle A., G. Theile: HDTV-Mehrkanalton: Untersuchungen zur Abbildungsqualität beim Einsatz zusätzlicher Mittenlautsprecher (HDTV-multichannel sound: investigations on the imaging quality with additional center speakers). In *16. Tonmeistertagung, VDT International Convention*, 208–218 (Verband deutscher Tonmeister, VDT, 1990)

Sippl F.: Stereo Sound for TV & Video. In *1st AES UK Conference* (1988) (Paper Number SWP-A1)

Sivian L. J., S. D. White: On Minimum Audible Sound Fields. J. Acoust. Soc. Am. **4**, 288–321 (1933)

Sivonen V. P.: Directional loudness and binaural summation for wideband and reverberant sounds. J. Acoust. Soc. Am. **121**, 2852–2861 (2007)

Snow W. B.: Effect of Arrival Time on Stereophonic Localization. J. Acoust. Soc. Am. **26**, 1071–1074 (1954)

Snow W. B.: Basic Principles of Stereophonic Sound. IRE Transactions on Audio **3**, 42–53 (1955)

Spikofski G., M. Fruhmann: Optimisation of Binaural Room Scanning (BRS): Considering inter-individual HRTF-characteristics. In *19th International AES Conference* (2001) (Paper Number 1907)

Spikofski G., G. Stoll, G. Theile: Neue Untersuchungsergebnisse zur Bestimmung von Kopfhörer-Übertragungsmaßen mit Hilfe der Sonden-Vergleichsmessung (New results on the determination of headphone magnitude transfer functions using the probe comparison measurement). In *Fortschritte der Akustik, DAGA '86*, 773–776 (DPG, Bad Honnef, 1986)

Sporer T., B. Klehs: Wave Field Synthesis in the Real World: Part 2 – In the Movie Theatre. In *116th AES Convention* (2004) (Convention Paper 6055)

Spors S.: *Active Listening Room Compensation for Spatial Sound Reproduction Systems*, PhD thesis, Friedrich-Alexander-Universität Erlangen-Nürnberg (2005)

Spors S., J. Ahrens: Comparison of Higher-Order Ambisonics and Wave Field Synthesis with Respect to Spatial Aliasing Artifacts. In *19th International Congress on Acoustics (ICA)* (2007)

Spors S., J. Ahrens: A Comparison of Wave Field Synthesis and Higher-Order Ambisonics with Respect to Physical Properties and Spatial Sampling. In *125th AES Convention* (2008) (Convention Paper 7556)

Spors S., R. Rabenstein, J. Ahrens: The Theory of Wave Field Synthesis Revisited. In *124th AES Convention* (2008) (Convention Paper 7358)

Start E. W.: *Direct sound enhancement by wave field synthesis*, PhD thesis, Technische Universiteit Delft (1997)

Steinberg J. C.: The Relation between the Loudness of a Sound and its Physical Stimulus. Phys. Rev. **26**, 507–523 (1925)

Steinberg J. C., W. B. Snow: Symposium on Wire Transmission of Symphonic Music and Its Reproduction in Auditory Perspective – Physical Factors. Bell System Technical Journal **13**, 245–258 (1934)

Steinberg Media Technologies GmbH: *ASIO 2.2 Audio Streaming Input Output – Development Kit, Interface Specification.* 2.2 Edition. Hamburg, Germany (2006)

Stemplinger I., M. Schiele, B. Meglić, H. Fastl: Einsilberverständlichkeit in unterschiedlichen Störgeräuschen für Deutsch, Ungarisch und Slowenisch (Intelligibility of monosyllables in different background noises for German, Hungarian, and Slovenian). In *Fortschritte der Akustik, DAGA '97*, 77–78 (Dt. Gesell. für Akustik e. V., Oldenburg, 1997)

Stevens S. S., H. Davis: *Hearing – Its Psychology and Physiology.* (John Wiley & Sons, New York, London, Sydney, 1938) (Seventh Printing 1966)

Stinson M. R.: The spatial distribution of sound pressure within scaled replicas of the human ear canal. J. Acoust. Soc. Am. **78**, 1596–1602 (1985)

Stinson M. R., G. A. Daigle: Comparison of an analytic horn equation approach and a boundary element method for the calculation of sound fields in the human ear canal. J. Acoust. Soc. Am. **118**, 2405–2411 (2005)

Stinson M. R., S. M. Khanna: Sound propagation in the ear canal and coupling to the eardrum, with measurements on model systems. J. Acoust. Soc. Am. **85**, 2481–2491 (1989)

Stinson M. R., B. W. Lawton: Specification of the geometry of the human ear canal for the prediction of sound-pressure level distribution. J. Acoust. Soc. Am. **85**, 2492–2503 (1989)

Stoll G., G. Theile: Gegenüberstellung von Lautstärke- und Sondenvergleichsmessungen zur Bestimmung des Kopfhörer-Übertragungsmaßes im Diffusen Schallfeld (Comparison of loudness and probe comparison measurements for determining the headphone magnitude transfer function in the diffuse sound field). In *Fortschritte der Akustik, DAGA '86*, 777–780 (DPG, Bad Honnef, 1986)

Strauss H.: Implementing Doppler Shifts for Virtual Auditory Environments. In *104th AES Convention* (1998) (Preprint 4687)

Stumpf C.: Differenztöne und Konsonanz (Difference tones and consonance). Zeitschr. f. Psychol. **39**, 269–283 (1905)

Sundaram S., C. Kyriakakis: Phantom Audio Sources with Vertically Seperated Speakers. In *119th AES Convention* (2005) (Convention Paper 6614)

Suzuki B. H.: DVD-Audio and DVD-Audio Recording Specifications. In *18th International Congress on Acoustics (ICA)*, III 2177–2178 (2004)

Takeuchi T., P. A. Nelson: Optimal source distribution for binaural synthesis over loudspeakers. J. Acoust. Soc. Am. **112**, 2786–2797 (2002)

Takeuchi T., P. A. Nelson: Subjective and Objective Evaluation of the Optimal Source Distribution for Virtual Acoustic Imaging. J. Audio Eng. Soc. **55**, 981–997 (2007)

Takeuchi T., P. A. Nelson, H. Hamada: Robustness to head misalignment of virtual sound imaging systems. J. Acoust. Soc. Am. **109**, 958–971 (2001)

Terhardt E.: Fourier Transformation of Time Signals: Conceptual Revision. Acustica **57**, 242–256 (1985)

Terhardt E.: *Akustische Kommunikation (Acoustic communication).* (Springer, Berlin, Heidelberg, New York, 1998)

Theile G.: Weshalb ist der Kammfilter-Effekt bei Summenlokalisation nicht hörbar? (Why is the comb-filter effect in summing localization inaudible?). In *11. Tonmeistertagung, VDT International Convention*, 200–214 (Verband deutscher Tonmeister, VDT, 1978)

Theile G.: *Über die Lokalisation im überlagerten Schallfeld (On the localization in the superimposed soundfield)*, PhD thesis, Technische Universität Berlin (1980)

Theile G.: Zur Theorie der optimalen Wiedergabe von stereofonen Signalen über Lautsprecher und Kopfhörer (On the theory of optimum reproduction of stereophonic signals by way of loudspeakers and headsets). Rundfunktech. Mitteilungen **25**, 155–170 (1981)

Theile G.: Sind "Klangfarbe" und "Lautstärke" vollständig determiniert durch das Schalldruckpegel-Spektrum am Trommelfell? (Are "sound color" and "loudness" completely determined by the sound pressure level spectrum at the eardrum?) In *Fortschritte der Akustik, DAGA '84*, 747–752 (DPG, Bad Honnef, 1984)

Theile G.: Beurteilungskriterien für Kopfhörer unter Berücksichtigung verschiedener Anwendungsbereiche (Evaluation criteria for headphones considering different fields of application). In *NTG-Fachtagung Hörrundfunk*, 290–301 (VDE-Verlag, 1985)

Theile G.: On the Standardization of the Frequency Response of High-Quality Studio Headphones. J. Audio Eng. Soc. **34**, 956–969 (1986)

Theile G.: On the Performance of Two-Channel and Multi-Channel Stereophony. In *88th AES Convention* (1990) (Preprint 2887)

Theile G.: On the Naturalness of Two-Channel Stereo Sound. J. Audio Eng. Soc. **39**, 761–767 (1991)

Theile G.: Natural 5.1 Music Recording Based on Psychoacoustic Principles. In *19th International AES Conference* (2001) (Paper Number 1904)

Theile G., G. Plenge: Localization of Lateral Phantom Sources. J. Audio Eng. Soc. **25**, 196–200 (1977)

Theile G., H. Wittek: Principles in Surround Recordings with Height. In *130th AES Convention* (2011) (Convention Paper 8403)

Theile G., M. Link, G. Stoll: Low-bit rate coding of high quality audio signals. In *82nd AES Convention* (1987) (Preprint 2432)

Theile G., H. Wittek, M. Reisinger: Wellenfeldsynthese-Verfahren: Ein Weg für neue Möglichkeiten der räumlichen Tongestaltung (Wave field synthesis methods: A way for new possibilities of spatial sound design). In *22. Tonmeistertagung, VDT International Convention* (Verband deutscher Tonmeister, VDT, 2002)

Thurlow W. R., P. S. Runge: Effect of Induced Head Movements on Localization of Direction of Sounds. J. Acoust. Soc. Am. **42**, 480–488 (1967)

Toole F. E.: The Acoustics and Psychoacoustics of Loudspeakers and Rooms – The Stereo Past and the Multichannel Future. In *109th AES Convention* (2000) (Preprint 5201)

Toole F. E., B. M. Sayers: Lateralization Judgments and the Nature of Binaural Acoustic Images. J. Acoust. Soc. Am. **37**, 319–324 (1965)

Torger A., A. Farina: Real-Time Partitioned Convolution for Ambiophonics Surround Sound. In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 195–198 (2001)

Traunmüller H.: Analytical expressions for the tonotopic sensory scale. J. Acoust. Soc. Am. **88**, 97–100 (1990)

Treeby B. E., J. Pan, R. M. Paurobally: The effect of hair on auditory localization cues. J. Acoust. Soc. Am. **122**, 3586–3597 (2007a)

Treeby B. E., J. Pan, R. M. Paurobally: An experimental study of the acoustic impedance characteristics of human hair. J. Acoust. Soc. Am. **122**, 2107–2117 (2007b)

Trevino J., T. Okamoto, Y. Iwaya, Y. Suzuki: High order Ambisonic decoding method for irregular loudspeaker arrays. In *20th International Congress on Acoustics (ICA)* (2010)

Usher J., W. L. Martens: Perceived Naturalness of Speech Sounds Presented Using Personalized Versus Non-Personalized HRTFs. In *13th International Conference on Auditory Display (ICAD)* (2007)

Verheijen E. N. G.: *Sound Reproduction by Wave Field Synthesis*, PhD thesis, Technische Universiteit Delft (1997)

Villchur E.: Free-Field Calibration of Earphones. J. Acoust. Soc. Am. **46**, 1527–1534 (1969)

Vogel P.: *Application of Wave Field Synthesis in Room Acoustics*, PhD thesis, Technische Universiteit Delft (1993)

Völk F.: Externalization in data-based Binaural Synthesis: Effects of Impulse Response Length. In *International Conference on Acoustics NAG/DAGA 2009*, 1075–1078 (Dt. Gesell. für Akustik e. V., Berlin, 2009)

Völk F.: Messtechnische Verifizierung eines datenbasierten binauralen Synthesesystems (Metrological verification of a data based binaural synthesis system). In *Fortschritte der Akustik, DAGA 2010*, 1049–1050 (Dt. Gesell. für Akustik e. V., Berlin, 2010a)

Völk F.: Psychoakustische Experimente zur Distanz mittels Wellenfeldsynthese erzeugter Hörereignisse (Psychoacoustic experiments on the distance of auditory events created by wave field synthesis). In *Fortschritte der Akustik, DAGA 2010*, 1065–1066 (Dt. Gesell. für Akustik e. V., Berlin, 2010b)

Völk F.: Inter- and Intra-Individual Variability in Blocked Auditory Canal Transfer Functions of Three Circum-Aural Headphones. In *131st AES Convention* (2011a) (Convention Paper 8465)

Völk F.: System Theory of Binaural Synthesis. In *131st AES Convention* (2011b) (Convention Paper 8568)

Völk F.: Headphone Selection for Binaural Synthesis with Blocked Auditory Canal Recording. In *132nd AES Convention* (2012a) (Convention Paper 8677)

Völk F.: *Chapter* Audio playback for auditory quality evaluations – Requirements, possibilities, and the impact on applications. In *Motor- und Aggregate-Akustik IV (Motor and aggregate acoustics IV)*, ed. by H. Tschöke, W. Henze, T. Luft, 31–47 (Expert Verlag, Renningen, 2012b)

Völk F., H. Fastl: Advantages of binaural room synthesis for research and fitting of hearing aids, cochlear implants, electro-acoustical stimulation, and combined systems. In *20th International Congress on Acoustics (ICA)* (2010)

Völk F., H. Fastl: Locating the Missing 6 dB by Loudness Calibration of Binaural Synthesis. In *131st AES Convention* (2011a) (Convention Paper 8488)

Völk F., H. Fastl: Richtungsunterschiedsschwellen (Minimum Audible Angles) für ein zirkulares Wellenfeldsynthesesystem in reflexionsbehafteter Umgebung (Minimum audible angles for a circular wave field synthesis system in a reverberant environment). In *Fortschritte der Akustik, DAGA 2011*, 945–946 (Dt. Gesell. für Akustik e. V., Berlin, 2011b)

Völk F., H. Fastl: Wave Field Synthesis with Primary Source Correction: Theory, Simulation Results, and Comparison to Earlier Approaches. In *133rd AES Convention* (2012) (Convention Paper 8717)

Völk F., E. Faccinelli, H. Fastl: Überlegungen zu Möglichkeiten und Grenzen virtueller Wellenfeldsynthese (Considerations of possibilities and limitations of virtual wave field synthesis). In *Fortschritte der Akustik, DAGA 2010*, 1069–1070 (Dt. Gesell. für Akustik e. V., Berlin, 2010a)

Völk F., M. A. Fintoc, H. Fastl: Minimum Audible Angles with Dynamic Binaural Synthesis and Bilateral Noise Vocoder: Effects of the Channel Distribution. In *Fortschritte der Akustik, DAGA 2012*, 773–774 (Dt. Gesell. für Akustik e. V., Berlin, 2012a)

Völk F., A. Gottschalk, H. Fastl: Evaluation of Binaural Synthesis by Minimum Audible Angles. In *Fortschritte der Akustik, DAGA 2012*, 321–322 (Dt. Gesell. für Akustik e. V., Berlin, 2012b)

Völk F., F. Heinemann, H. Fastl: Externalization in binaural synthesis: effects of recording environment and measurement procedure. In *Acoustics '08*, 6419–6424 (2008a)

Völk F., J. Konradl, H. Fastl: Simulation of wave field synthesis. In *Acoustics '08*, 1165–1170 (2008b)

Völk F., C. Landsiedel, H. Fastl: Auditory Adapted Exponential Transfer Function Smoothing (AAS). In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 49–52 (2011a)

Völk F., F. Lindner, H. Fastl: Primary Source Correction (PSC) in Wave Field Synthesis. In *ICSA 2011 – International Conference on Spatial Audio* (Verband deutscher Tonmeister, VDT, 2011b)

Völk F., U. Mühlbauer, H. Fastl: Minimum Audible Distance (MAD) by the Example of Wave Field Synthesis. In *Fortschritte der Akustik, DAGA 2012*, 319–320 (Dt. Gesell. für Akustik e. V., Berlin, 2012c)

Völk F., T. Musialik, H. Fastl: Crosstalk Cancellation between Phantom Sources. In *126th AES Convention* (2009a) (Convention Paper 7722)

Völk F., T. Riesenweber, H. Fastl: Ein Algorithmus zur Vorhersage des für transparente Auralisierung nutzbaren Dynamikbereichs rauschbehafteter Impulsantworten (An algorithm for the prediction of the dynamic range available for transparent auralization in noisy impulse responses). In *Fortschritte der Akustik, DAGA 2011*, 315–316 (Dt. Gesell. für Akustik e. V., Berlin, 2011c)

Völk F., M. Schmidhuber, H. Fastl: Influence of the ventriloquism effect on minimum audible angles assessed with wave field synthesis and intensity panning. In *20th International Congress on Acoustics (ICA)* (2010b)

Völk F., M. Straubinger, H. Fastl: Psychoacoustical experiments on loudness perception in wave field synthesis. In *20th International Congress on Acoustics (ICA)* (2010c)

Völk F., A. Dunstmair, T. Riesenweber, H. Fastl: Bedingungen für gleichlaute Schalle aus Kopfhörern und Lautsprechern (Requirements for equally loud sounds from headphones and loudspeakers). In *Fortschritte der Akustik, DAGA 2011*, 825–826 (Dt. Gesell. für Akustik e. V., Berlin, 2011d)

Völk F., S. Kerber, H. Fastl, S. Reifinger: Design und Realisierung von virtueller Akustik für ein Augmented-Reality-Labor (Design and implementation of virtual acoustics for an augmented reality lab). In *Fortschritte der Akustik, DAGA 2007*, 673–674 (Dt. Gesell. für Akustik e. V., Berlin, 2007)

Völk F., M. Straubinger, L. Roalter, H. Fastl: Measurement of Head Related Impulse Responses for Psychoacoustic Research. In *International Conference on Acoustics NAG/DAGA 2009*, 164–167 (Dt. Gesell. für Akustik e. V., Berlin, 2009b)

von Békésy G.: A new audiometer. Acta Oto-laryngol. **35**, 411–422 (1947)

von Békésy G.: On the Resonance Curve and the Decay Period at Various Points on the Cochlear Partition. J. Acoust. Soc. Am. **21**, 245–254 (1949) von Bismarck G.: Timbre of steady sounds: Scaling of sharpness. In *7th International Congress on Acoustics (ICA)*, Vol. 3, 637–640 (1971)

von Bismarck G.: Sharpness as an attribute of the timbre of steady sounds. Acustica **30**, 159–172 (1974)

von Helmholtz H.: Ueber einige Gesetze der Vertheilung elektrischer Ströme in körperlichen Leitern mit Anwendung auf die thierisch-elektrischen Versuche (On some laws concerning the distribution of electrical currents in conductors with applications to experiments on animal electricity). Annalen der Physik **165**, 211–233 (1853)

Vorländer M.: Acoustic load on the ear caused by headphones. J. Acoust. Soc. Am. **107**, 2082–2088 (2000)

Vorländer M.: *Auralization – Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality.* (Springer, Berlin, Heidelberg, 2008)

Wallach H.: On Sound Localization. J. Acoust. Soc. Am. **10**, 270–274 (1939)

Wallach H.: The Role of Head Movements and Vestibular and Visual Cues in Sound Localization. J. Exp. Psychol. **27**, 339–368 (1940)

Wallach H., E. B. Newman, M. R. Rosenzweig: The precedence effect in sound localization. Am. J. Psychol. **62**, 315–336 (1949)

Wallerus H.: Richtungsauflösungsvermögen des Gehörs für Sinustöne mit interauralen Pegelunterschieden (Directional resolution of the hearing system for sinusoidal tones with interaural level differences). In *Fortschritte der Akustik, DAGA '76*, 589–592 (VDI, Düsseldorf, 1976)

Ward D. B.: Joint Least Squares Optimization for Robust Acoustic Crosstalk Cancellation. IEEE Transactions on Speech and Audio Processing **8**, 211–215 (2000)

Ward D. B., G. W. Elko: Effect of Loudspeaker Position on the Robustness of Acoustic Crosstalk Cancellation. IEEE Signal Processing Letters **6**, 106–108 (1999)

Warncke H.: Die Grundlagen der raumbezüglichen stereophonischen Übertragung im Tonfilm (The basics of room-related stereophonic transmission in the movie with sound). Akust. Z. **6**, 174–188 (1941)

Webers J.: *Handbuch der Tonstudiotechnik (Handbook of recording studio technique).* 9th Edition (Franzis, Poing, 2007)

Wefers F., S. Pelzer, M. Vorländer: Recording natural sound sources and implementing them in virtual acoustic scenes. In *Fortschritte der Akustik, DAGA 2011*, 325–326 (Dt. Gesell. für Akustik e. V., Berlin, 2011)

Weingartner B. A.: Vergleich von objektiven Kopfhörermessungen an Kupplern mit Sondenmessungen am Ohr (Comparison between objective headphone measurements on couplers and measurements with probe microphone on the ear). In *2nd AES Central Europe Convention* (1972) (Paper Number C15)

Wente E. C., A. L. Thuras: Symposium on Wire Transmission of Symphonic Music and Its Reproduction in Auditory Perspective – Loud Speakers and Microphones. Bell System Technical Journal **13**, 259–277 (1934)

Wenzel E. M.: The Relative Contribution of Interaural Time and Magnitude Cues to Dynamic Sound Localization. In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 80–83 (1995)

Wenzel E. M., J. D. Miller, J. S. Abel: A software-based system for interactive spatial sound synthesis. In *6th International Conference on Auditory Display (ICAD)* (2000)

Wenzel E. M., F. L. Wightman, D. J. Kistler: Localization with non-individualized virtual acoustic display cues. In *SIGCHI Conference on Human Factors in Computing Systems*, 351–359 (1991)

Wenzel E. M., M. Arruda, D. J. Kistler, F. L. Wightman: Localization using nonindividualized head-related transfer functions. J. Acoust. Soc. Am. **94**, 111–123 (1993)

Wiegrebe L., K. Krumbholz, W. Schmid, S. Schmidt: Detektion transienter Signale – Ein Maß für Intensitätsintegration im Gehör? (Detection of transient signals – A measure for intensity integration in the hearing system?). In *Fortschritte der Akustik, DAGA '96*, 332–333 (Dt. Gesell. für Akustik e. V., Oldenburg, 1996)

Wightman F. L., D. J. Kistler: Headphone simulation of free-field listening. I: Stimulus synthesis. J. Acoust. Soc. Am. **85**, 858–867 (1989a)

Wightman F. L., D. J. Kistler: Headphone simulation of free-field listening. II: Psychophysical validation. J. Acoust. Soc. Am. **85**, 868–878 (1989b)

Wightman F. L., D. J. Kistler: The Perceptual Relevance of Individual Differences in Head-Related Transfer Functions. Acustica – Acta Acustica **82**, S 92 (1996)

Wightman F. L., D. J. Kistler: Monaural sound localization revisited. J. Acoust. Soc. Am. **101**, 1050–1063 (1997)

Wightman F. L., D. J. Kistler: Resolution of front-back ambiguity in spatial hearing by listener and source movement. J. Acoust. Soc. Am. **105**, 2841–2853 (1999)

Wightman F. L., D. J. Kistler: Measurement and Validation of Human HRTFs for Use in Hearing Research. Acta Acustica united with Acustica **91**, 429–439 (2005)

Wightman F. L., D. J. Kistler, M. Arruda: Perceptual consequences of engineering compromises in synthesis of virtual auditory objects. J. Acoust. Soc. Am. **92**, 2332 (1992)

Wilkens H.: Kopfbezügliche Stereophonie – ein Hilfsmittel für Vergleich und Beurteilung verschiedener Raumeindrücke (Head-Related Stereophony – An Aid for the Comparison and Critical Examination of Different Room Effects). Acustica **26**, 213–221 (1972)

Winkler H.: Das Sehen beim Hören (Seeing when hearing). In *Fortschritte der Akustik, DAGA '92*, 181–184 (DPG, Bad Honnef, 1992)

Wittek H.: *Perceptual differences between wavefield synthesis and stereophony*, PhD thesis, University of Surrey (2007)

Wittek H., T. Augustin: Räumliche Wahrnehmung von Wellenfeldsynthese: Der Einfluss von Alias-Effekten auf die Klangfarbe (Spatial perception of wave field synthesis: The influence of alias artifacts on the sound color). In *Fortschritte der Akustik, DAGA '05*, 199–200 (Dt. Gesell. für Akustik e. V., Berlin, 2005)

Wittek H., F. Rumsey, G. Theile: On the sound colour properties of wavefield synthesis and stereo. In *123rd AES Convention* (2007a) (Convention Paper 7167)

Wittek H., F. Rumsey, G. Theile: Perceptual Enhancement of Wavefield Synthesis by Stereophonic Means. J. Audio Eng. Soc. **55**, 723–751 (2007b)

Xie B., T. Zhang: The Audibility of Spectral Detail of Head-Related Transfer Functions at High Frequency. Acta Acustica united with Acustica **96**, 328–339 (2010)

Xu S., Z. Li, G. Salvendy: Individualization of Head-Related Transfer Function for Three-Dimensional Virtual Auditory Display: A Review. In *12th International Conference on Human-Computer Interaction (HCI)*, 397–407 (2007)

Yang W. Y., T. G. Chang, I. H. Song, Y. S. Cho, J. Heo, W. G. Jeon, J. W. Lee, J. K. Kim: *Signals and Systems with MATLAB®*. (Springer, Berlin, Heidelberg, 2009)

Yoshida M., A. Kudo, H. Hokari, S. Shimada: Impact of equalizing ear canal transfer function on out-of-head sound localization. In *123rd AES Convention* (2007) (Convention Paper 7229)

Zahorik P.: Assessing auditory distance perception using virtual acoustics. J. Acoust. Soc. Am. **111**, 1832–1846 (2002)

Zahorik P., E. Brandewie, V. P. Sivonen: Spatial Hearing in Reverberant Rooms and Effects of Prior Listening Exposure. In *International Workshop on the Principles and Applications of Spatial Hearing (IWPASH)* (2009)

Zahorik P., D. S. Brungart, A. W. Bronkhorst: Auditory Distance Perception in Humans: A Summary of Past and Present Research. Acta Acustica united with Acustica **91**, 409–420 (2005)

Zarouchas T., J. Mourjopoulos: Modeling perceptual effects of reverberation on stereophonic sound reproduction in rooms. J. Acoust. Soc. Am. **126**, 229–242 (2009)

Zhong X.-L., Y. Liu, N. Xiang, B.-S. Xie: Errors in the measurements of individual headphone-to-ear-canal transfer function. In *Fortschritte der Akustik, DAGA 2010*, 607–608 (Dt. Gesell. für Akustik e. V., Berlin, 2010)

Zieglmeier W., G. Theile: Darstellung seitlicher Schallquellen bei Anwendung des 3/2-Stereo Formates (Imaging of Lateral Sources Using the 3/2-Stereo Format). In *19. Tonmeistertagung, VDT International Convention*, 159–169 (Verband deutscher Tonmeister, VDT, 1996)

Zollner M.: Einfluß von Stativen und Halterungen auf den Mikrophonfrequenzgang (Influence of Stands and Mountings on the Frequency Response of Microphones). Acustica **51**, 268–272 (1982)

Zollner M.: Methodisch bedingte Fehler bei binauralen Hörversuchen (Methodically induced errors in binaural listening experiments). In *Fortschritte der Akustik, DAGA '95*, 779–782 (Dt. Gesell. für Akustik e. V., Oldenburg, 1995)

Zollner M., E. Zwicker: *Elektroakustik (Electroacoustics).* 3$^{\text{rd}}$ Edition (Springer, Berlin, 1993)

Zölzer U.: *Digitale Audiosignalverarbeitung (Digital audio signal processing).* 3$^{\text{rd}}$ Edition (B. G. Teubner, Stuttgart, Leipzig, Wiesbaden, 2005)

Zuckerwar A. J., R. Meredith: Low-frequency absorption of sound in air. J. Acoust. Soc. Am. **78**, 946–955 (1985)

Zwicker E.: On the Loudness of Continuous Noises. J. Acoust. Soc. Am. **28**, 764 (1956)

Zwicker E.: Über psychologische und methodische Grundlagen der Lautheit (On psychological and methodical basics of loudness). Acustica **8**, 237–258 (1958)

Zwicker E.: Der akustische Eingangswiderstand des äußeren Ohres (The acoustic input impedance of the external ear). Nachrichtentechnische Zeitung **8**, 397–404 (1961a)

Zwicker E.: Subdivision of the Audible Frequency Range into Critical Bands (Frequenzgruppen). J. Acoust. Soc. Am. **33**, 248 (1961b)

Zwicker E.: Temporal Effects in Simultaneous Masking and Loudness. J. Acoust. Soc. Am. **38**, 132–141 (1965)

Zwicker E.: Der Einfluss der Zeitstruktur verdeckender Klänge auf die Mithörschwelle (The influence of the temporal masker structure on the masking pattern). In *Fortschritte der Akustik, DAGA '75*, 323–326 (Physik, Weinheim, 1975)

Zwicker E.: Die Abbildung der Schalldruck-Zeitfunktion im Mithörschwellen-Periodenmuster (The representation of the sound-pressure time function in the masking-period pattern). Acustica **34**, 189–199 (1976a)

Zwicker E.: Masking period patterns of harmonic complex tones. J. Acoust. Soc. Am. **60**, 429–439 (1976b)

Zwicker E.: A Model for Predicting Masking-Period Patterns. Biol. Cybernetics **23**, 49–60 (1976c)

Zwicker E.: Über die Phasenbeziehungen zwischen Schalldruck und Erregung (On the phase relation of sound pressure and excitation). In *Fortschritte der Akustik, DAGA '76*, 605–608 (VDI, Düsseldorf, 1976d)

Zwicker E.: Procedure for calculating loudness of temporally variable sounds. J. Acoust. Soc. Am. **62**, 675–682 (1977)

Zwicker E.: A Model Describing Nonlinearities in Hearing by Active Processes with Saturation at 40 dB. Biol. Cybernetics **35**, 243–250 (1979)

Zwicker E.: Dependence of post-masking on masker duration and its relation to temporal effects in loudness. J. Acoust. Soc. Am. **75**, 219–223 (1984)

Zwicker E.: Das Innenohr als aktives schallverarbeitendes und schallaussendendes System (The inner ear as an active sound processing and sound emitting system). In *Fortschritte der Akustik, DAGA '85*, 29–44 (DPG, Bad Honnef, 1985)

Zwicker E.: A hardware cochlear nonlinear preprocessing model with active feedback. J. Acoust. Soc. Am. **80**, 146–153 (1986)

Zwicker E.: Procedure for calculating partially masked loudness based on ISO 532 B. In *Inter-Noise '87*, 1021–1024 (1987)

Zwicker E.: Unmasked and partially masked loudness in musical dynamics. J. Acoust. Soc. Am. **85**, S 140 (1989)

Zwicker E., G. Bubel: Einfluß nichtlinearer Effekte auf die Frequenzselektivität des Gehörs (Influence of the Nonlinearity of the Hearing System on its Frequency Selectivity). Acustica **38**, 67–71 (1977)

Zwicker E., H. Fastl: On the Development of the Critical Band. J. Acoust. Soc. Am. **52**, 699–702 (1972)

Zwicker E., R. Feldtkeller: Über die Lautstärke von gleichförmigen Geräuschen (On the loudness of uniform sounds). Acustica **5**, 303–316 (1955)

Zwicker E., R. Feldtkeller: *Das Ohr als Nachrichtenempfänger (The ear as a communication receiver).* 2$^{\text{nd}}$ Edition (S. Hirzel Verlag, Stuttgart, 1967)

Zwicker E., R. Feldtkeller: *The ear as a communication receiver.* (Translated from German by H. Müsch, S. Buus, and M. Florentine, published for the Acoustical Society of America through the American Institute of Physics, AIP Press, Woodbury, New York, 1999)

Zwicker E., G. Gässler: Die Eignung des dynamischen Kopfhörers zur Untersuchung frequenzmodulierter Töne (The suitability of a dynamic earphone for the measurement of FM-tones). Acustica **Akustische Beihefte**, AB 134–139 (1952)

Zwicker E., G. B. Henning: On the effect of interaural phase differences on loudness. Hearing Research **53**, 141–152 (1991)

Zwicker E., D. Maiwald: Über das Freifeldübertragungsmaß des Kopfhörers DT 48 (On the free-field response of the earphone DT 48). Acustica **13**, 181–182 (1963)

Zwicker E., E. Terhardt: Analytical expressions for critical-band rate and critical bandwidth as a function of frequency. J. Acoust. Soc. Am. **68**, 1523–1525 (1980)

Zwicker E., U. T. Zwicker: Audio Engineering and Psychoacoustics: Matching Signals to the Final Receiver, the Human Auditory System. J. Audio Eng. Soc. **39**, 115–126 (1991)

Zwicker E., H. Fastl, C. Dallmayr: BASIC-Program for calculating the loudness of sounds from their 1/3-oct band spectra according to ISO 532 B. Acustica **55**, 63–67 (1984)

Zwicker E., G. Flottorp, S. S. Stevens: Critical Band Width in Loudness Summation. J. Acoust. Soc. Am. **29**, 548–557 (1957)

Zwicker U. T.: Psychoacoustics as the basis for modern audio signal data compression. J. Acoust. Soc. Am. **107**, 2875 (2000)