Technische Universität Berlin

Fakultät I. Fachgebiet Audiokommunikation

Magisterarbeit

Individual headphone compensation for binaural synthesis



vorgelegt von Fabian Brinkmann

berlin

Abgabe: 21.8.2011

Erstgutachter Prof. Dr. Stefan Weinzierl Zweitgutachter Alexander Lindau, M.A.

Eidesstattliche Erklärung

Hiermit versichere ich gegenüber der Fakultät I der Technischen Universität Berlin, dass ich die vorliegende Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. Alle Ausführungen, die anderen veröffentlichten oder nicht veröffentlichten Schriften wörtlich oder sinngemäß entnommen wurden, habe ich kenntlich gemacht. Die Arbeit hat in gleicher oder ähnlicher Fassung noch keiner anderen Prüfungsbehörde vorgelegen.

Berlin, den 20. August 2011 _____

Fabian Brinkmann

Acknowledgement

I am much obliged to Alexander Lindau, Zora Schärer, Frank Schultz, Michael Horn, Sebastian Roos, Matthias Geier & Sascha Spors (T-Labs), Klaus Heinz (Adam Audio), Alfred Stirnemann (Phonak), Joachim Feldmann, Edgar Berdahl, Sabine Unger, Walter Brinkmann, Raffael Töngens, Hans-Joachim Maempel & Stefan Weinzierl.

Abstract

In this work, the influence of individual, generic and non-individual headphone compensation on the perceived quality of a binaural simulation was examined. Therefore, the binaural simulation, originating from recordings made with a hat and torso simulator (HATS), was compared to the corresponding real sound field radiated by a loudspeaker in two quality evaluating listening tests. Individual recordings were not considered. For this setup, nonindividual headphone compensation turned out to be perceptively best suited, if it was based on headphone transfer functions measured on the same HATS. A true non-individual compensation, based on transfer functions from a third person however, was perceptively inferior to all other compensation approaches. Further, it was shown, that a subwoofer can be integrated into the binaural simulation, to enhance the playback of signals with strong low-frequency components, and that a minimum phase compensation, which might decrease the latency of the simulation, could be used without considerable perceptual degradation. For measuring individual binaural signals, a measurement instrument has been developed and evaluated in the course of this work.

Zusammenfassung

In der vorliegenden Arbeit wurde der Einfluss einer individuellen, generischen und nichtindividuellen Kopfhörerentzerrung auf die perzeptive Qualität der binauralen Simulation untersucht. Dafür wurde die binaurale Simulation, basierend auf Messungen mit einem Kopfund Torso-Simulator in zwei qualitätsbewertenden Hörversuchen mit dem korrespondierenden realen Schallfeld, abgestrahlt durch einen Lautsprecher, verglichen. Für den untersuchten Fall wurde die nicht-individuelle Entzerrung, wenn sie auf Grundlage von Kopf- und Torso-Simulator Kopfhörerübertragungsfunktionen durchgeführt wurde, gegenüber der individuellen und generischen Entzerrung bevorzugt. Eine wirklich nicht-individuelle Entzerrung, basierend auf Übertragungsfunktionen einer dritten Person, schnitt dagegen schlechter als alle anderen Kompensationsmethoden ab. Des weiteren konnte gezeigt werden, dass die binaurale Simulation um einen Subwoofer ergänzt werden kann, der die Wiedergabe von Signalen mit ausgeprägten tieffrequenten Anteilen ermöglicht, und dass eine die Latenz der Simulation verringernde minimalphasige Entzerrung zu keiner nennenswerten perzeptiven Beeinträchtigung führt. Für die Messung individueller binauraler Signale wurde im Zuge der Arbeit ein Messinstrument angefertigt und evaluiert.

Contents

Glo	ossary		iii
1.	Motivati	on and scope	1
2.	State of	research	3
	2.1. Bi	naural hearing	3
	2.2. Bi	naural synthesis	6
	2.2	2.1. Binaural signals	6
	2.2	2.2. A compensation filter for binaural synthesis	8
	2.2	2.3. Auralization	10
	2.2	2.4. Evaluation	12
	2.3. Re	cording binaural signals of human subjects	13
	2.4. He	adphone compensation	17
	2.4	4.1. LMS compensation – Linear phase	18
	2.4	4.2. LMS compensation – Minimum phase	22
	2.5. Th	e influence of binaural signals	23
	2.5	5.1. HRTFs and BRTFs	23
	2.5	5.2. HPTFs and headphone compensation filter	25
	2.5	5.3. Other Influences	27
	2.6. Ch	apter summary	28
3.	Physical	evaluation	31
	3.1. De	eveloping a measuring instrument	31
	3.2	1.1. Anatomy of the human ear canal	32
	3.2	1.2. Microphones for binaural recordings	35
	3.2	1.3. Microphone measurement and inversion	36
	3.2	1.4. Prototyping	42
	3.1	1.5. Crafting	44
	3.2	1.6. Physical evaluation of the measuring instrument	46
	3.2. Ev	aluation of headphone compensation	48
	3.2	2.1. In situ HPTF measurement	48
	3.2	2.2. Headphone compensation	51
	3.2	2.3. Compensation results	54
	3.2	2.4. Auditory modeling of compensation results	59

	3.3.	Chapter summary	61
4.	Perceptual Evaluation		
	4.1.	Listening Test I	63
		4.1.1. Design, sample and measure	64
		4.1.2. Setup and validation	67
		4.1.3. Procedure	72
		4.1.4. Analysis and Results	73
		4.1.5. Discussion	76
	4.2.	Listening Test II	80
		4.2.1. Possible compensation improvements	81
		4.2.2. Sample	85
		4.2.3. Results	86
		4.2.4. Discussion	88
	4.3.	Chapter summary	89
5	Conc	lusion	91
0.	5.1.	Summary	91
	5.2	Perspectives	92
	0.2.		/2
Bił	oliogra	aphy	95
Lis	t of F	igures	103
Lis	t of T	ables	105
A.	Spec	ification of technical equipment	106
Β.	Perce	eptual Evaluation	117
C.	Othe	r	135
D.	Elect	ronic documentation	136

Glossary

ANOVA	analysis of variance (ANOVA).		
BRIR BRTF	binaural room impulse response. binaural room transfer function.		
FABIAN	Fast and Automatic Binaural Impulse response AcquvisitioN. A head and torso simulator developed by Lindau (2006).		
FEC	free air equivalent coupling.		
HATS	head and torso simulator.		
hp	headphone.		
HPTF	headphone transfer function.		
HRIR	head related impulse response.		
HRTF	head related transfer function.		
ILD	interaural level difference.		
IPD	interaural phase difference.		
ITD	interaural time difference.		
LMS	least mean square.		
LTI	linear and time invariant.		
PRECISE	ECISE precisely repeatable acquisition of individual headphone transfer functions (PRECISE). Silicon ear moulds with flush-cast microphones.		

Chapter 1

Motivation and scope

Over the last decades, realization and evaluation of virtual acoustic environments (VAEs) has been widely discussed in literature. By means of dynamic binaural synthesis, VAEs can be created, recording the signals arriving at a listeners ears and reproducing them through headphones or loudspeakers. The areas of application for such systems are broad, and different demands can be made towards the naturalness of the simulation.

In teleconferencing or as a tool in the early acoustical design stages of rooms, a plausible simulation, that means "a simulation in agreement with the listener's expectation towards an equivalent real acoustic event" (Lindau and Weinzierl, 2011, p. 1), could be sufficient. Other applications, as the evaluation of the quality of loudspeakers or rooms, where absolute judgements are required, could demand an authentic simulation. In this case a "perceived identity between simulation and reality" (Lindau and Weinzierl, 2011, p. 1) is desired. The main influence factor on the quality of binaural synthesis, and therefore on its authenticity, are the recordings used for reproducing the ear signals. These are typically non-individual, not recorded using the listeners head and ears but a head and torso simulator (HATS). Lindau and Weinzierl (2011) and Schärer (2008) showed, that using such non-individual recordings, a plausible but not authentic simulation can be yield¹.

If binaural signals are reproduced via headphones, their influence has to be compensated in order to ensure, that the recorded signals can be reproduced unaltered at a listeners ears (Møller, 1992). The influence of two different headphones and seven compensation methods on the perceived quality of the binaural simulation was examined by Schärer (2008) in a previous study. Using recordings from a HATS and non-individual headphone compensation based on headphone transfer functions (HPTFs) from the same HATS, the so called high pass regulated least mean square (LMS) inversion turned out to be perceptively best suited.

However it seems not to be clear, wether or not individual headphone compensation has to be preferred. Pralong and Carlile (1996) showed that both, individual and nonindividual binaural recordings can be reproduced nearly perfect at a listeners ears, if an individual headphone compensation is applied. On the other hand, a precise reproduction of non-individual recordings must not necessarily be the best choice regarding authenticity, i.e. the "precieved identity" of a binaurally simulated and the corresponding real sound

¹Strictly spoken, authenticity was not tested in Schärer (2008), but as the simulation was distinguishable, it can be said not to be authentic.

field. Surprisingly, a systematic investigation of this effect has not been done yet. In most cases, investigations are either focused on a physical description of the effect (Pralong and Carlile, 1996) or cover only partial aspects of binaural hearing like localization (Møller et al., 1996b; Minnaar et al., 2001). Therefore the headphone compensation is the main scope of the present study. Besides individual and non-individual headphone compensation, a generic one, based on an average HPTFs of many subjects, was considered as well. Thus, the research question was: Which headphone compensation (individual, generic, non-individual) is perceptually best suited if non-individual BRIRs are used? Before being able to assess this question, a reliable measurement tool for the acquisition of individual HPTFs had to be developed, which allows a fast in situ measurement.

Besides this main research question, two additional aspects arouse from the study conducted by Schärer (2008). First, the headphone compensation filter can be designed as a minimum or linear phase system. Both possibilities exhibit potentially audible advantages and disadvantages regarding the temporal behavior of the simulation, which were examined. Second, subjects from Schärer reported, that the simulation had a "poor bass" compared to the real sound field, though both were equalized to yield identical frequency responses between 50 Hz and 21 kHz. It was argued that the headphones are unable to simulate the influence of low frequencies on the human auditory and tactile system. As this would be a major drawback regarding an authentic simulation, it was examined wether or not a subwoofer could be employed to improve low frequency reproduction.

It has to be noted, that although the investigation of the mentioned aspects possibly lead to an improvement of the binaural simulation in view of authenticity, this work was not aimed at creating such an authentic simulation. Rather, based on the work of Schärer, the effects of headphone compensation and the deployment of a subwoofer on the perceived quality of the binaural simulation were investigated. Therefore, a listening test was conducted, where a simulated and the corresponding real sound field could be compared directly, and qualitative judgements on the similarity were obtained .

For reasons of feasibility, the present study was restricted to the use of non-individual binaural room impulse responses (BRIRs), recorded with a HATS. Nevertheless, as a reliable measuring instrument is needed for the acquisition of individual binaural recordings, the present study is part of the effort of creating an authentic binaural reference system, that is being undertaken by the Audio Communication Group at Technical University Berlin in conjunction with the DFG research unit Simulation and Evaluation of Acoustical Environments (SEACEN)¹.

¹See http://www.seacen.tu-berlin.de/menue/seacen/parameter/en/ (Last checked: July 2011)

Chapter 2

State of research

Before examining the effect of individual headphone compensation on binaural synthesis, one has to look at the mechanisms behind binaural hearing $\Rightarrow 2.1$ Binaural hearing. The sensitivity of the auditory system towards changes in the signal arriving at the two ears, for example evoked by the displacement of a sound source or by the manipulation of level and phase information, can be used to establish quality criteria towards the binaural simulation. These criteria then can be used for introducing a framework for binaural simulation in theory and practice $\Rightarrow 2.2$ Binaural synthesis, with a particular focus on the headphone compensation $\Rightarrow 2.4$ Headphone compensation. Regarding the desired measuring instrument for individual HPTFs, methods for recording binaural signals of human subjects are reviewed with respect to reliability and feasibility $\Rightarrow 2.3$ Recording binaural signals of human subjects. Finally, the variance in binaural signals is discussed, to give an estimate of their influence on the binaural signals.

2.1 Binaural hearing

Using the head-related spherical coordinate system (Fig. 2.1), the position of a sound source with respect to the listener can be described by azimuth (left-right) and elevation (up-down) angles ϕ and θ and the distance r between its center and the sound source. The center is defined as the middle of the interaural axis, that passes through the entrances of the ear canals. The point $p(\phi, \theta, r) = (0^{\circ}, 0^{\circ}, 1 \text{ m})$ then denotes a sound source about 1 meter in front of the listener on the intersection of horizontal and median plane. In the following, a brief overview of binaural hearing will be given, based on the comprehensive analysis done by Blauert (1997).

In localizing a sound source, mainly three mechanisms are involved. First and second, the sound from a source to the left or right of the median plane, i.e. with an azimuth angle which is not zero, will arrive earlier at the ear closer to the source causing an interaural time difference (ITD) and will have a bigger level at that ear resulting in an interaural level difference (ILD)¹. The sensitivity of binaural hearing towards ITDs and ILDs has been investigated in listening tests presenting systematically manipulated stimuli through headphones. It is described by the so called lateralization blur, which is defined as the smallest change in a parameter that

¹These main cues were first investigated by Strutt (Lord Rayleigh), establishing the *Duplex theory of hearing*



Figure 2.1. Head-related spherical coordinate system (from Vorländer, 2008).

leads to a perceivable source displacement. Lateralization, as opposed to localization, is used in conjunction with listening through headphones where in head localization occurs. Localization, on the other hand, is used, if a sound source is perceived to be located outside the head.

The smallest ITD values have been reported by Klemm $(1920)^1$. He discovered a lateralization blur of $2 - 10 \ \mu s$ using click signals. Mills (1958) investigated the sensitivity of the human auditory system towards phase differences. He reports just noticeable interaural phase differences (IPDs) of $2 - 4^{\circ}$ for sine signals in the range of 150-1000 Hz. The smallest lateralization blur found for ILDs is 0.6 dB for a 2 kHz sine signal given by Ford (1942)². Furthermore it was noted that a diffuse sound source or even two sound sources can be perceived when only the ITD is manipulated or when the ILD is strongly varying with frequency (Whitworth and Jeffress, 1961; Toole and Sayers, 1965)³. In any case, the sensitivity towards interaural time and level differences greatly varies with level and signal.

Since isolated manipulation of ITDs and ILDs is only possible with headphones, it is not clear how much it affects binaural hearing in natural environments. In this case the sensitivity of binaural hearing to source displacement is described by means of localization blur, defined as the smallest perceivable displacement of a sound source. The best localization blur can be observed in front of a listener ($\phi = \theta = 0^\circ$). It is $\pm 0.75^\circ$ in the horizontal and $\pm 9^\circ$ in the median plane (Klemm, 1920; Wettschurek, 1970)⁴.

The third mechanism is involved in localization in the median plane. In the absence of

¹citet after Blauert (1997, p. 153)

²ibid. p. 161

³ibid. p. 160

⁴ibid. p. 39 and p. 44

Localization	n blur	Later	Lateralization blur	
Horiz. plane	$\pm 0.75^{\circ}$	ITD	$2 - 10 \ \mu s$	
Median plane	$\pm9^{\circ}$	IPD	$2-4^{\circ}$	
		ILD	0.6 dB	

Table 2.1. Localization and lateralization blur.

interaural time and level differences, spectral coloration caused by the human legs, torso, shoulders, head and especially the pinna, delivers useful information. Blauert (1997, p. 69) describes the pinna as "[...] a system of acoustical resonators. The degree to which individual resonances of this system are exited depends on the direction and distance of the sound source." Other physical phenomena that lead to spectral coloration are reflection, shadowing, dispersion, diffraction and interference. Depending on the elevation of a sound source certain frequencies are thus attenuated or amplified, which supports localization. Spectral coloration is also used in detecting the distance of a sound source. Both, distance detection and localization in the median plane, strongly depend on the signal, and work best with familiar broad band sounds (Blauert, 1997, Chap. 2.3).

According to Blauert (1997, Chap. 2.5.2) the influence of bone-conducted sound on binaural hearing can be neglected, but other senses can very well have an influence on it. The perceived location of a sound source for example can vary, wether or not it is also presented visually. Usually the perceived location of the auditory event is dominated by the visual location, if they differ from each other to a limited extend. In addition to that, head movements can help in localizing a sound source. Tactile sensations evoked by high pressure levels and low frequencies do not support localization, but certainly are a considerable attribute of a sound source.

So far, only issues concerning listening in free field conditions were discussed. However, other attributes, related to listening in rooms, where reflections from the room boundaries add information to the sound signals, clearly contribute to binaural hearing, too. They can be described by means of room acoustical parameters, like the early decay time (EDT) or speech intelligibility measures. These parameters are determined with pressure or pressure gradient receivers. Binaural measures as the interaural cross-correlation function (IACF), that might be used as an estimate for apparent source width (ASW) or listener envelopment (LEV), are still in the minority (Ahnert and Tennhardt, 2008; Vorländer, 2008, Chap. 6.4). Although such binaural measures might be of interest, as they hold spectral and temporal information induced by head and pinna, a detailed description of such measures and their influence on binaural hearing lies beyond the scope of this study and is part of ongoing research.

2.2 Binaural synthesis

If all information evaluated in binaural hearing was inherent to the signals at both eardrums, it could be simulated by recording and reproducing those signals at a listeners ear. Møller (1992) demonstrated that this is in fact true and further showed that either headphones or loudspeakers can be used for reproducing binaural recordings. Loudspeaker reproduction involves two compensation filters, one for cross-talk cancelation, that usually has to be adapted to the listeners head orientation and its position within the room, and a second for compensating the influence of head and pinna. Two possible solutions to this problem are discussed by Lentz et al. (2005) and Menzel et al. (2005). Although the present study is focused on headphone reproduction of binaural recordings, part of the results may hold true for loudspeaker reproduction as well, as the second filter mentioned above is comparable to the headphone compensation, described in the following.

2.2.1 Binaural signals

The free-field sound transmission from a point in space to the ear canal is described by the head related transfer function (HRTF). By applying the inverse fourier transform, the head related impulse response (HRIR) is obtained, which is the time domain equivalent of the HRTF. Both representations contain all spatial information inherent to the source and consequentially depend on source azimuth, elevation and distance and of course on the subject being measured on. Middlebrooks et al. (1989) and Hammershøi and Møller (1991)¹ showed, that the sound transmission from a point a few millimeters outside the ear canal to the ear drum is already independent of direction. Hence, in a broader sense, the transfer function from a point in space to a point within or at the entrance to the open or blocked ear canal is also understood as HRTF. Following the nomenclature of Møller (1992) (see Fig. 2.2 and Tab. 2.2), it can be written as

$$HRTF_{ed}(\phi, \, \theta, \, r, \, \omega, \, subject, \, ear) = \frac{P_4}{P_1}, \tag{2.1}$$

$$HRTF_{oe}(\phi, \theta, r, \omega, subject, ear) = \frac{P_3}{P_1},$$
(2.2)

or

$$HRTF_{be}(\phi, \theta, r, \omega, subject, ear) = \frac{P_2}{P_1},$$
(2.3)

¹Or Hammershøi and Møller (1996) for a more detailed discussion

where *ed*, *oe* and *be* specify the measuring positions ear drum, entrance to open and blocked ear canal respectively. Note that capital letters denote signals in the frequency and lower case letters in the time domain, and that *p* is short for p(t), as well as *P* is short for $P(\omega)$. Abbreviations however are mostly written in capital letters, no matter if they refer to time or frequency domain signals.



Figure 2.2. Sound transmission through the external ear. (a) anatomic and (b) electroacoustic model; (c) anatomic and (d) electroacoustic model with headphone as source. Taken from Møller (1992). For a nomenclature see Tab. 2.2.

Sound transmission in rooms is described by the binaural room transfer function (BRTF) and BRIR respectively. The difference compared to free-field listening is that there is not only sound arriving directly from the source, but also reflections from the room boundaries that reach the listeners ears from various azimuth and elevation angles. Assuming acoustic sound transmission to be linear and time invariant (LTI)¹, the BRTF can be described as a superposition of HRTFs

$$BRTF = \sum_{i=1}^{N} e^{-jwt_i} w_i \ HRTF_i \ , \tag{2.4}$$

¹For a definition of LTI systems see Oppenheim et al. (2004)

F	Free field Headpho	
p_1	sound pressure at middle of head without listener	p_1
p_2	sound pressure at blocked ear canal	p_5
p_3	sound pressure at entrance to ear canal	p_6
p_4	sound pressure at eardrum	p_7

Table 2.2. Sound pressures at the external ear for free field and
headphone sound transmission (see Fig. 2.2).

where w_i is a weighting coefficient depending on the wall material of the room, the entire path length from source to listener, the air absorption and source directivity. e^{-jwt_i} is the delay introduced by the travel path (Vorländer, 2008, Chap. 15.2.2). The BRTF then depends on the source position with respect to the listener, the position and orientation of the listener within the room, the HRTFs of the listener and of course on the room itself.

The sound transmission from a headphone to the eardrum, or in a broader sense to a point within or at the entrance to the open or blocked ear canal, is described by the HPTF which is given by

$$HPTF_{ed}(\omega, hp, ear) = \frac{P_7}{E_{hp}},$$
(2.5)

$$HPTF_{oe}(\omega, hp, ear) = \frac{P_6}{E_{hp}},$$
(2.6)

and

$$HPTF_{be}(\omega, hp, ear) = \frac{P_5}{E_{hp}}.$$
(2.7)

Again following the nomenclature introduced by Møller (1992), e_{hp} is the voltage at the headphone (hp) terminals. As binaural signals are recorded with microphones, the sound pressures P_i can be substituted with E_{mic}/M , where E_{mic} is the voltage at the microphone terminals and M its frequency response.

2.2.2 A compensation filter for binaural synthesis

Transducers involved in recording and reproducing binaural signals, namely the measuring microphones and loudspeakers as well as the headphones, introduce unwanted spectral coloration and phase distortion (Fig. 2.3). Their influence can be compensated by means of a digital filter, as demonstrated by the thorough analysis of Møller (1992).

Looking at the role of the headphones first, their influence can be compensated by filtering



Figure 2.3. Recording and reproducing binaural signals. H_c : compensation filter; $M_{1, 2}$: Microphones. Transducers are colored red.

the binaural recordings with the inverse of the HPTF (Møller, 1992, derived from Eq. 30, 34, 38).

$$H_{c,HPTF,ed}(\omega) = \frac{1}{HPTF_{ed}}$$
(2.8)

$$H_{c, HPTF, oe}(\omega) = \frac{1}{HPTF_{oe}}$$
(2.9)

$$H_{c, HPTF, be}(\omega) = \frac{1}{HPTF_{be}} \cdot \frac{Z_{ear \ canal} + Z_{hp}}{Z_{ear \ canal} + Z_{radiation}}$$
(2.10)

The additional term in Eq. 2.10 indicates that the situation is more complicated when recording binaural signals at the entrance to the blocked ear canal. The term becomes unity, if the acoustical impedances at the exit of the ear canal Z_{hp} and $Z_{radiation}$ (see Fig. 2.2) are the same for free-field and headphone listening conditions. Møller et al. (1995b) investigated which headphones meet this so called free air equivalent coupling (FEC) criterion by measuring pressure division ratios

$$PDR = \frac{P_3/P_2}{P_6/P_5} = \frac{Z_{ear\ canal} + Z_{hp}}{Z_{ear\ canal} + Z_{radiation}}$$
(2.11)

of 14 headphones. Four out of the 14 headphones, two extraaural and two circumaural, met the criterion with a tolerance of ± 2 dB, if average levels were considered. The measurements were report to be valid up to 7 kHz, above, small displacements of the microphone lead to errors, that could not be compensated. In theory the PDR could be included in the compensation filter $H_{c, hp, be}$, in practice this is barely done, since measuring is time-consuming. Thus Eq. 2.8-2.10 can be simplified to

$$H_{c, HPTF}(\omega) = \frac{1}{HPTF},$$
(2.12)

if FEC headphones are used.

Including the frequency responses $M_1(\omega)$ and $M_2(\omega)$ (see Fig. 2.3) of microphones used for binaural recordings, the compensation filter becomes

$$H_{c}(\omega) = \frac{1}{M_{1} \left[P_{5;6;7}/E_{hp} \right]}$$

= $\frac{1}{M_{1} \left[(E_{2}/M_{2})/E_{hp} \right]}$
= $\frac{1}{M_{1}} \cdot M_{2} \frac{E_{hp}}{E_{2}}$ (2.13)

where E_2 is the voltage at the microphone terminals. Equation 2.13 looks straight forward, but the nature of HPTFs makes its calculation a rather complex topic, as will be discussed in Chap. 2.4.

The compensation filter does not account for the loudspeaker, that is used for binaural recordings. In semi-diffuse environments, its influence can only be compensated by modeling the directivity of the source that should be reproduced, e.g. a violin or a singer (Lindau, 2006). Measuring and modeling source directivities can be done with spherical microphone and loudspeaker arrays. See Meyer (2008) for an overview of classical instruments directivities, and Zotter (2009) for a theoretical discussion of loudspeaker and microphone arrays.

2.2.3 Auralization

Following Kleiner et al. (1993, p. 861), "Auralization is the process of rendering audible, by physical or mathematical modeling, the sound field of a source in a space, in such a way as to simulate the binaural listening experience at a given position in the modeled space." Regarding binaural synthesis, this is achieved by convoluting anechoic audio material with HRTFs or BRTFs and playback through headphones. In the following, the auralization framework used throughout this study, as depicted in Fig. 2.4, will be described briefly. It has been developed at the Audio Communication Group at Technical University Berlin.

For the acquisition of BRIR/HRIR datasets the FABIAN (Fast and Automatic Binaural Impulse response AcquisitioN) HATS as described in Lindau (2006) is used. FABIAN's head and outer ears are casts of a human head (Moldrzyk, 2002). DPA 4060 electret condenser microphones are mounted at the entrance to the blocked ear canal¹. Head and torso are connected by a neck joint, allowing head movements with an accuracy of $\pm 0.02^{\circ}$ and within a wide range ($-90^{\circ} \le \phi \le 90^{\circ}$, $-45^{\circ} \le \theta \le 90^{\circ}$). Torso and arms are abstractions derived from average male and female anthropometric data. BRIR datasets including multiple sources can be measured with a software written in Matlab©, that automatically steps through a

¹See Appendix A for technical specifications of equipment used in this study.



Figure 2.4. Auralization using dynamic binaural synthesis. Red lines represent audio and grey lines control signals. Dashed lines denote offline processing.

predefined grid of head and torso orientations.

An anechoic audio stream is then convolved with a pair of BRIRs¹. Convolution is done in the frequency domain with a unix based software described in Lindau et al. (2007). For efficient computing, it splits the convolution process in two parts (a) the early reflections, dynamically chosen from the BRIR pair according to the head orientation, which is tracked using a Polhemus FASTRAK© and (b) statistical reverberation chosen from a predefined but fixed BRIR pair. The transition between early reflections and statistical reverberation is marked by the so called mixing time. The perceptual mixing time was systematically investigated by Lindau et al. (2010b). However the results were not available when this study was carried out. Instead, results from Lindau et al. (2007) were used, who examined a relatively wet room (V=10000 m³, T_{30, 1 kHz}=2 sec.). Subsequent to the convolution of the anechoic audio stream with the BRIRs, the compensation Filter H_c is applied using the same convolution engine.

Lastly, the ITDs, that have been extracted from the BRIR dataset before auralization, are reinserted with a variable length delay-line. The ITDs are scaled to fit the listener using the inter tragus distance as the predictor. Among other advantages, this methods reduces the system latency and improves localization of simulated sources (Lindau et al., 2010a).

The signal processing and the technical equipment used for auralization introduces a latency, defined by the time span between a head movement and the output of the corresponding audio signal. Schärer (2008, p. 115) reports this latency on average to be 73.8 ms, however it can be reduced, as shown in Lindau (2009b).

¹For an introduction to digital signal processing including convolution see Oppenheim et al. (2004)

2.2.4 Evaluation

This chapter gives a brief overview of studies, that evaluated the auralization framework described above. The most sensitive way of assessing the quality of a binaural simulation, is to directly compare it to a real sound field. This as been done by Moldrzyk et al. (2005) as well as Lindau et al. (2007). Both conducted listening tests using the AB comparison test paradigm. Without a reference to the real sound field, the majority of subjects was not able to reliably detect the simulation. Nevertheless, subjects reported that differences between simulation and reality were audible.

While having a reference to the real sound field, subjects from Schärer (2008) were asked to detect the simulation and rate its difference compared to the real sound field. The simulation was clearly distinguishable from the corresponding real sound field by attributes listed in Tab. 2.3.

Moldrzyk et al. (2005)	Lindau et al. (2007)	Schärer (2008)		
tone color	spectral differences	high frequency ringing		
reverberation	source localization	timbre		
localization	reverberant energy	poor bass		
distance	energy on contralateral ear	spatiality		
	loudness	localization		
	latency	transients		
	tone change during head	source distance		
	movements	latency		
		loudness		
		naturalness		
Classification: Coloration (red), spatiality (yellow), localization (green), temporal behavior (blue),				

lassification: Coloration (red), spatiality (yellow), localization (green), temporal behavior (blue others (grey).

Table 2.3. Attributes allowing the distinction between real and simulated sound field.Sorted by relevance.

The results of the three mentioned studies are in good accordance to each other, suggesting that spectral attributes related to tone color, followed by attributes describing the spatial impression as well as localization, are most often used to distinguish the simulated from the real sound field. Differences in loudness were only mentioned occasionally and could as well interact with spectral differences. As will be discussed in Chap. 2.5, the reason for these differences are most likely the non-individual binaural recordings used for auralization. Another factor influencing the spectral coloration is the headphone compensation, which is subjected in the present study. Subjects from Schärer (2008) also reported, that the binaural

simulation had a "poor bass" compared to the real sound field. This indicates, that a) the headphones were not able to transmit low frequencies with the needed energy or b) tactile sensations evoked by low frequencies could not be reproduced (see Chap. 2.1)¹.

Furthermore, Schärer (2008) mentions, that the linear phase headphone compensation adds an additional delay to the simulation. This might increase the total system latency to a value higher than the threshold observed by Lindau (2009b), and may also influence the transient behavior (see Chap. 2.4)

2.3 Recording binaural signals of human subjects

As mentioned before, the sound transmission from a point about 6 mm outside the entrance of the ear canal to the ear drum is independent of direction. Hence binaural signals can be measured anywhere between these points. In general, measurement positions can be subdivided into positions at the open and at the blocked ear canal. In any case, a good repeatability of the microphone position is desired in order to minimize the error induced by the measurement method itself (Riederer, 1998).

For measuring at the open ear canal, mostly so called probe microphones are used. They typically consist of the probe tip, a flexible plastic tube, that is attached to a cavitiy enclosing a pressure transducer. The probe tip has a visible length of up to 7 cm and an outer diameter ranging from 1 - 2.5 mm. That way, probe microphones allow for the use of relatively large microphones, that would not fit the ear canal and for measuring inside the ear canal without or only negligible disturbance of the sound field. One problem that arises with measurements inside the ear canal is that the probe tip measures at a specific point, while the ear drum integrates over its cross-section (Møller, 1992). This however can be disregarded, considering the ear canal as a cylindrical tube with rigid boundaries, which holds true in good approximation (Blauert, 1997, p. 56). In this case, cross modes within the ear canal only exist above a cut-on frequency given by

$$f_1 = 0.59 \frac{c}{d}$$
(2.14)

or close to inconsistencies in the tube, such as the entrance to the ear canal and its termination at the ear drum (Möser, 2009, Chap. 6). Assuming an average diameter of d = 8 mm for the ear canal and a speed of sound of $c = 340 \frac{m}{s}$, Eq. 2.14 leads to a cut-on frequency of 25.1 kHz, that lies well outside the frequency range perceivable by the human auditory system. Furthermore, Middlebrooks et al. (1989) reported that the influence of cross modes

¹An inspection revealed, that this was probably caused by Schärer's test setup. So it remains unclear, wether or not the headphones are able two properly reproduce low frequencies.

appearing at inconsistencies on the sound field is minimal. However, within the ear canal longitudinal standing waves are present, and hence the sound pressure measured at the open ear canal almost exclusively depends on the length the probe tip is inserted into it.

Examples for probe microphones placed at the open ear canal are given in Fig. 2.5 (a) taken from Hammershøi and Møller (1996) and (b) and (c) from Wightman and Kistler (1989), who used custom earmolds to guide the probe tip and hold it in place. They deploy a custom build probe microphone, where the probe tip is attached to miniature electret condenser microphone (ETYMŌTIC RESEARCH). A similar method for microphone placement has been used by Pralong and Carlile (1994). Another option for holding the probe tip at a desired position would be fixing it with medical tape. The amount, by what it is inserted into the ear canal can be marked on the probe tip itself. That way, measurement positions are believed to be well reproducable.



(a) (b) (c) (c) (c) (d)

Figure 2.5. Examples for microphone positioning from various studies.

For measuring at the blocked ear canal both, probe and miniature microphones can be used. The latter were build since the mid 1970's and since then found widespread use, because they are usually cheaper than probe microphones while having a higher sensitivity and thus yielding a better signal to noise ratio. For blocking the ear canal, three methods are used across literature, that vary in the amount of blockage and the precision the microphone can be placed with.

The most common method is to block the ear with a compressible foam earplug, that is inserted into the ear canal and than extends to provide a good fit. Often, the cylindrical *E-A-R Classic*¹ is used, as depicted in Fig. 2.5 (d) and (e) taken from Møller et al. (1995b). The microphone is placed in a hole, cut or burned into the earplug. Though the *E-A-R Classic*, with a diameter of 13 mm and a length of 18 mm, seems to be quite large to fit the average ear canal, Riederer (2004b) is the only one mentioning problems regarding the size of the earplug. He reports difficulties with positioning the microphone and with the fit to small ear canals. The latter may lead to an invalid measuring position, if the earplug sticks out from the ear canal (see Fig. 2.5 (f) taken from Riederer (2004b)). It is also mentioned that, if the ear canals of a subject are to small, measurements could not be conducted using the *E-A-R Classic*.

The second possibility is to fill the entrance of the ear canal with moldable silicon and place the microphone on top. This method ensures a better blockage than using foam earplugs and a good hold of the microphone. But still, the microphone has to be repositioned if measuring on successive days. Because of the complex resonance structure at the blocked ear canal, displacements of the microphone of 1 - 2 mm may have quite a big influence on the resulting measurements (Riederer, 2004b,a, see Fig. 2.5 f).

The most sophisticated way for measuring at the blocked ear canal would be building custom earmolds with flush-cast microphones, as done for the *LISTEN* project² shown Fig. 2.5 (h). The advantage of this approach is clearly the well defined measuring position even for measurements conducted on different dates.

The main quality criterion regarding the measurement procedure of binaural signals is its repeatability that is mainly determined by the microphone positioning accuracy. However the investigation of this is somehow cumbersome, because other effects like the subjects position relative to the measuring equipment (loudspeaker or headphone) and small head movements also influence the results. Altough the topic of the present study was the measurement and compensation of HPTFs, the discussion on repeatability will be carried out on HRTFs. The reason for this is that the placement of headphones already induces great variability to the measurements, as will be seen in Chap. 2.5.

Møller et al. (1995b,c) investigated the effects of microphone positioning accuracy and

¹See http://www.aearo.com/pdf/hearing/Econopack.pdf (Last checked: May 2011)

²See http://recherche.ircam.fr/equipes/salles/listen/index.html (Last checked: May 2011) and Eckel (2001)

head movements in summation. Their results showed comparable variances for a series of measurements conducted at the blocked ear canal using probe and miniature microphones, suggesting that neither method is superior to the other. However they complained, that the housing of the probe microphone might have noticeably disturbed the sound field arriving at the ears.

Riederer (1998) tried to assess the repeatability at the blocked ear canal using the *E-A-R Classic* and moldable silicon. Three differently experienced experimenters inserted the earplugs. Five HRTFs were measured on one human subject for various incidents of sound and for each earplug. The results for the *E-A-R Classic* exhibited deviations of about ± 1 dB below 1 kHz, ± 2.5 dB below 5 kHz and ± 5 dB above 5 kHz. Moreover, deviations of up to ± 15 dB were seen in regions of spectral notches roughly above 6 kHz which is in good agreement with Algazi et al. (1999). The deviations were slightly worse for the less experienced experimenters. Different, though slightly better results, can be observed for blockage with moldable silicone. Riederer carefully controlled the position of the subject and reported an inacuracy of approximately 2 cm for his procedure, which equaled an azimuth offset of 1°. Despite this, it has to be assumed that the subject placement and its head movements influenced the measurements, and in a later study using a comparable setup, a magnitude of about ± 2 dB was given for this error (Riederer, 2004a). Regarding this, the error induced by positioning inaccuracy for measuring at the blocked ear canal would be negligible up to 5 kHz.

Very similar results can be observed in 10 repeated measurments carried out with a probe microphone at the open ear canal of one human subject (Wightman and Kistler, 1989). In this case, custom ear molds were used to keep the probe tip in its position, but comparable results should be obtainable by fixing the probe tip with medical tape.

Besides the repeatability of the microphone position, deviations from the desired rigid blockage of the ear canal are another source of errors. Foam ear plugs, which increasingly behave absorbing at high frequencies, or leakage of silicon earplugs could evoke such deviations. According to Mellert, the membranes of condenser microphones can be regarded as rigid, whereas probe microphones can not (Blauert et al., 1978). This would advise the use of miniature electret condenser microphones for measuring at the blocked ear canal.

In summary, it can be stated that a similar precision can be obtained measuring at any point at the open or blocked ear canal and by using any method. The error induced by the microphone placement is small in general, but nevertheless could affect the binaural simulation at high frequencies, because of the sensitivity of binaural hearing towards ILDs (see Tab. 2.1).

2.4 Headphone compensation

This chapter deals with the calculation of a headphone compensation filter based on measured HPTFs. The aim is to compensate (equalize, linearize) the binaural signal transmission chain $H(\omega)$, that is shown in Fig. 2.3, within the range given by a predefined target bandpass $D(\omega)$:

$$H_{eq}(\omega) = H(\omega) \cdot H_c(\omega) = D(\omega), \qquad (2.15)$$

or

$$h_{eq}(t) = h(t) * h_c(t) = d(t)$$
 (2.16)

respectively. H_c is the compensation filter

$$H_c(\omega) = \frac{1}{M_1} \cdot M_2 \frac{E_{hp}}{E_2} ,$$

as introduced in Eq. 2.13 and H_{eq} the resulting compensation (Schärer, 2008)¹. It is assumed, that the frequency responses $M_1(\omega)$ and $M_2(\omega)$ of the microphones are known and their inversion is relatively simple, as descripted in Chap. 3.1.3. Hence, this chapter focuses on the calculation of

$$\frac{E_{hp}}{E_2}$$

,

of which the reciprocal can be measured and than has to be inverted.

In a previous study, Schärer (2008) investigated the perceptual suitability of different filter inversion algorithms, and the so called high pass regulated LMS inversion showed the best results. Schärer mentioned the following general problems, that arose with the inversion process:

First, a complete compensation may cause excessive gains at frequencies close to 0 Hz or the Nyquist frequency. This can be avoided by a well designed target bandpass. Second, a direct inversion leads to an unstable filter, because HPTFs are known not to have minimum phase frequency responses (Minnaar et al., 1999) and third, an exact inversion is unwanted. Repositioning of the headphones causes a considerable variance in HPTFs, and thus an exact inversion may lead to audible ringing artifacts (see Chap. 2.5.2). The second and third point have to be accessed by the filter inversion algorithm, given in the following and summing up the work of Schärer.

¹An english, but shorter version is given by Schärer and Lindau (2009)

2.4.1 LMS compensation – Linear phase

The target bandpass, as depictet in Fig. 2.6, was designed with the firl function contained in the Signal Processing ToolboxTM of Matlab©, using a Kaiser window with a sidelobe attenuation of 60 dB. Schärer measured HPTFs of seven headphones, and regarding their frequency responses, -6 dB cutt-off frequencies of 50 Hz and 21 kHz have been chosen. To avoid potentially audible group delay distortions, the phase of the filter was chosen to be linear. Schärer denotes, that the filter could as well be designed as a minimum phase



Figure 2.6. Target bandpass used by Schärer (2008). Magnitude response (left), windowed time signal and kaiser window (top right) and group delay (bottom right).

system. But since minimum phase filters would introduce unwanted and audible group delay distortions, they were excluded from further considerations. The disadvantage of the linear phase response is the introduced delay, which is half the length of the filters impulse response.

In all, Schärer included seven different inversion methods in a listening test to assess their percetual suitability, always using the described bandpass as target function. The so called high-pass regulated LMS inversion has shown to be the perceptively best suited of the tested methods. This approach minimizes the energy of the error between compensation result and target function. The amount of compensation, i.e. the amount of gain that is applied by the compensation filter, can be limited with a regularization function and an overall regularization weight. On this account, excessive gains of the inversion filter are avoided. The LMS algorithm can be implemented in the time or frequency domain (Kirkeby and Nelson, 1999; Kirkeby et al., 1998). Both methods were tested by Schärer, and no significant differences could be observed between them¹. Because the calculation in the time

¹ANOVA, tested for small effects, significane level α = 0.05, power 0.8

domain is less efficiently, the frequency domain calculation was preferred. It is given by

$$H'_{c}(\omega) = \frac{D(\omega)HPTF^{*}(\omega)}{|HPTF(\omega)|^{2} + \beta|B(\omega)|^{2}} , \qquad (2.17)$$

where $HPTF^*(\omega)$ is the complex conjugate of $HPTF(\omega)$, and β and $B(\omega)$ are the regularization weight and function, respectively. A derivation of Eq. 2.17 can be found in Norcross et al. (2006). The effort of the regularization is limited in the passband of $B(\omega)$ and for high values of β . As can be seen from the nominator of Eq. 2.17, only the energy of $B(\omega)$ is considered, and thus the phase can be chosen arbitrarily without influencing the result. This also holds true for the time domain calculation (Schärer, 2008, p. 44).

The compensation filter $H_c(\omega)$ can then be achieved by applying the following three steps: (a) Measuring the HPTF, as defined in Eq. 2.5 - 2.7, and compensate it for M_2 :

$$HPTF = \frac{E_2}{E_{hp}} \cdot \frac{1}{M_2}$$

(b) Calculate the headphone compensation filter after Eq. 2.17:

$$H'_{c}(\omega) = \frac{D(\omega)HPTF^{*}(\omega)}{|HPTF(\omega)|^{2} + \beta|B(\omega)|^{2}}$$

(c) Multiply the headphone compensation filter with the inverse of M_1 :

$$H_c = H_c' \frac{1}{M_1}$$

The separation into different steps is necessary because the inversion of the HPTF, given by Eq. 2.17, is a non linear process. Step (c) could as well be excluded from the calculation of H_c , if the BRTFs or HRTFs used for auralization are compensated for M_1 .

As regularization function, a second order shelve with a low frequency gain of -20 dB and a mid-gain frequency of about 3 kHz was chosen by Schärer (see Fig. 2.7, red line). By this means, the effort of the compensation is limited at high frequencies, where high Q notches, whose center frequencies and gains are likely to vary for successive measurements, appear in the HPTFs. Thus an exact compensation of these notches is not wanted and may lead to excessive gains in the compensated HPTF, causing audible ringing artifacts (see Chap. 2.5.2). The mid-gain frequency has been adjusted by hand to ensure that the passband of the shelve filter starts at the point, where the first notch appears in the HPTF. The effect of the regularization weight β has been examined, and a value of 0.4 has been chosen. This assured a good trade-of between high overshoots, that occured for smaller β vales, and remaining deep notches in the compensation result, occurring for larger values. Furthermore, even double notches can evoke if to large values are used (Norcross et al., 2006).

The last parameter of the filter design is the filter length. Here, a compromise between a desired high frequency resolution and the additional delay introduced by the filter has to be made. Regarding this, Schärer chose a length of N=2049 Samples. Assuming a sampling frequency of $f_s = 44.1$ kHz, this leads to a frequency resolution of

$$\Delta f = \frac{fs}{N} = 21.5 \text{ Hz}$$

and a delay of

$$t = \frac{N}{2} = 23.2 \ ms$$
 .

For achieving a good slope steepness at the lower band edge given by $\triangle f$, the filter length should be about two or three times the cycle duration of $\triangle f$. Or the other way around, a good slope steepness can only be obtained above the threefold of $\triangle f$, in this case 64.5 Hz (Müller, 1999, p. 72).

Compensation results from Schärer (2008), obtained by the described procedure, are shown in Fig. 2.7. The average of ten HPTFs measured on FABIAN with *STAX SRS 2050 II* headphones was used to calculate H_c . The headphones have been repositioned between successive measurements. The averaging causes a smoothing of the frequency response and is another way of preventing an exact inversion, which was for example suggested by Kulkarni and Colburn (2000).

The magnitude responses resulting from applying H_c to ten HPTFs are shown as grey lines in the top of Fig. 2.7. The target bandpass is given in black and the regularization function in red. Noteworthy deviations from the target function can be found at two points: at frequencies below 150 Hz and above 3 kHz. At low frequencies, say between 20 Hz and 60 Hz, the compensated HPTFs are up to 5 dB below the target function. As discussed above, this is caused by the filter length, which is to short to achieve a proper slope steepness. Between 60 Hz and 150 Hz, deviations from the target function become symmetrical and are caused by variability in the HPTFs induced by leakage (see. Chap. 2.5.2). They lie within a range of ± 2 dB. Above 3 kHz the regularization causes the deviations to be asymmetrical in the range of -15 dB to +2.5 dB.

In the middle of Fig. 2.7 the group delay of an exemplary compensated HPTF is shown for left and right channel. Provided that no regularization is applied and a sufficient filter length



Figure 2.7. Compensation results from Schärer (2008).

is used, the linear-phase compensation should yield a constant group delay. Deviations from this constant group delay of 23.2 ms are seen below 100 Hz. In this range, the monaural group delay distortions exceeds 0.5 ms, a threshold given by Blauert and Laws (1978)¹ and therefore might be audible. Schärer (2008) notes that a higher filter order would solve this problem.

The phase difference between left and right channel is shown at the bottom of Fig. 2.7 for one compensated HPTF. At low frequencies, phase differences of about 5° can be observed, which exceeds the threshold given in Tab. 2.1. However, if this difference is perceivable with any other signals than pure tones, is questionable.

2.4.2 LMS compensation – Minimum phase

The advantage of a minimum phase filter design would be, that the criteria given by the target bandpass could be achieved with a lower filter order (Schärer, 2008, p. 34). It's disadvantage is the introduced group delay distortion, as mentioned before. In addition, two drawbacks of the linear phase solution are overcome: the additional delay introduced by the filter, and its symmetrical impulse response, where a considerable fraction of the energy appears before the main impulse. The latter could lead to audible preringing in the auralization (Norcross et al., 2006).

The algorithm to obtain a linear phase equalization, as described above, was modified by Norcross et al. (2006), and in this case Eq. 2.17 changes to

$$H'_{c}(\omega) = \frac{HPTF^{*}A(\omega)}{|HPTF|^{2}}.$$
(2.18)

The regularization function and weight, as well as desired phase response are contained in $A(\omega)$, given by

$$A(\omega) = A'(\omega) e^{j\phi(\omega)}, \qquad (2.19)$$

where $A'(\omega)$ holds the regularization:

$$A'(\omega) = \frac{1}{1 + \beta \frac{|B(\omega)|^2}{|HPTF(\omega)|^2}} .$$
(2.20)

The phase term in Eq. 2.19 can be chosen arbitrarily, and Norcross used the Hilbert transform to obtain a minimum phase response

$$\phi_{min}(\omega) = -\operatorname{imag}(\operatorname{Hilbert}(\ln(|A'(\omega)|))).$$
(2.21)

¹Citet after Schärer (2008, p. 81)

The target bandpass $D(\omega)$ can be applied by multiplying $A(\omega)$ with its absolute value. The compensation of $H(\omega)$ can then be computed, applying the three steps, listed for the linear phase compensation. Informal listening tests conducted by Norcross showed better results for the minimum, then for the linear phase equalization.

2.5 The influence of binaural signals on the binaural simulation

As seen in Chap. 2.2.4, the binaural simulation was clearly distinguishable from a corresponding real sound field, if non-individual recordings were used and a reference to the real sound field was given. Intuitively, using individual recordings should increase the quality. Regarding dynamic binaural synthesis, this would be time consuming and therefore hardly feasible if (a) a large number of sources or head orientations shall be simulated, (b) recordings are needed from many subjects or (c) a ready to use consumer auralization framework is desired. For these reasons, the drawbacks that arise from using non-individual, or said the other way around, the advantages from using individual recordings will be discussed in the following. This can be done in two ways, both taken into account: First by looking at the perceptual influences and second by examining the binaural transfer functions introduced in Chap. 2.2.1.

2.5.1 HRTFs and BRTFs

Many studies compared static binaural simulation to a real sound field. The most systematic and comprehensive work has been carried out by Møller and colleagues. They compared localization performance yield by binaural simulation and under natural listening conditions. 19 incidents of sound were considered, represented by 19 loudspeakers. Subjects were sitting in a standard listening room according to IEC 268-13 (1985), with the loudspeakers placed around them. Stimuli were presented through loudspeakers and headphones in separate sessions and subjects were asked to identify the loudspeaker that radiated the sound. Individual headphone compensation was applied throughout all experiments discussed in this section. Binaural recordings were made at the blocked ear canal.

Møller et al. (1996b) showed that localization performance under natural listening conditions is comparable to that using individual recordings. When non-individual recordings of human subjects are used, a significant increase of localization errors in the median plane, including front-back confusion, and errors in the detection of distance was observed. A significant influence on localization in the horizontal plane could not be found, indicating that ITD- and ILD-cues remain, even with non-individual recordings. Since spectral features play an important role for distance detection and localization in the median plane (see Chap. 2.1), the results of Møller implicitly support those from Schärer (2008), Lindau et al. (2007) and Moldrzyk et al. (2005). They reported spectral differences to be most prominent, when comparing real and simulated sound fields.

Further, Møller et al. (1996a) reported, that localization performance also depends on the non-individual recordings that are used. Carefully selected, errors in the median plane can be reduced to an amount comparable, but still differing from that using individual recordings. In other studies artificial heads and HATS have been evaluated. In general, localization performance was similar, but slightly worse than the average reported for the use of non-individual recordings made on human subjects (Møller et al., 1999; Minnaar et al., 2001).

In summary, it can be said that spectral differences dominate, if non-individual recordings are compared to individual recordings, or to a real sound field. An analysis of binaural transfer functions helps to quantify these differences. In principle, this analysis should be carried out on BRTFs, since the studies from Møller and colleagues were conducted in a reverberant environment. However BRTF are barely reported in literature, and because they can be seen as superpositions of HRTFs (see Eq. 2.4), the analysis of HRTFs should yield comparable results.



Figure 2.8. HRTFs of 43 human subjects left ears. Frontal (left) and contralateral (right) sound incidence in horizontal plane taken from the CIPIC HRTF database. One exemplary subject highlighted for clarity.

HRTFs of 43 human subjects taken from the *CIPIC HRTF database* are depicted in Fig. 2.8. (Algazi et al., 2001). For both incidents of sound, a common structure can be observed up to about 5 kHz. Apart from some outliers, deviations in this range rise from ± 2.5 dB to ± 5 dB. Above 5 kHz, characteristic notches and peaks emerge, whose frequencies, gains and Q's vary between subjects. In this range, deviations reach ± 20 dB. Notches and peaks are caused by pinna resonances of whom some have been identified by Shaw and Teranish (1968) and Shaw (1998) for measurements at the open and blocked ear canal, respectively. Hence, high frequency variation in HRTFs among subjects can possibly be explained by varying pinna
shapes.

Similar observations can be made for other incidents of sound, and results from other studies are in good accordance to those described above. A summary of HRTF studies can for example be found in Blauert (1997, p. 81) and Møller et al. (1995c, p.301).

Regarding the spectral differences in HRTFs, results from listening test mentioned in this section and in Chap. 2.2.4 can be well explained. Further, it is very unlikely, that an authentic simulation can be obtained by using non-individual HRTFs or BRTFs.

2.5.2 HPTFs and headphone compensation filter

Only few studies evaluated the perceptual influence of individual and non-individual headphone compensation on binaural synthesis. In most cases it was assumed, that individual compensation yields the best results, even when non-individual HRTFs or BRTFs are used. Findings from Møller et al. (1996a) seem to support this assumption. They reported slightly but significantly worse localization performance in the median plane for non-individual compared to individual headphone compensation in connection with non-individual BRTFs. In this case, the non-individual headphone compensation has been calculated on the basis of HPTFs averaged over several subjects.

As can be seen from Fig. 2.9, inter individual variations, i.e. differences between subjects, in HPTFs strongly depend on the headphone that is being used. HPTFs for each of the three headphones show a common structure up to 6 kHz (Møller et al., 1995b). The lowest deviations in this range can be seen for the circumaural *STAX SR Lambda professional*. They slowly rise from ± 1 dB to about ± 3 dB, whereas deviations from ± 3 dB to ± 5 dB occur with the extraaural *AKG K-1000*. For the supraaural *Sony MDR-102* headphones the biggest deviations of ± 5 dB can be seen. In this case, the high variations can be explained by leakage that is caused by poor seal between ear and headphone cushion (Dillon, 1977). Above 6 kHz, variations reach ± 15 dB for all of the three headphones. However a common structure is still seen for the circumaural and extraaural, but not for the supraaural transfer functions. If mean HPTFs of the three headphones are compared visually, a common structure among them can hardly be found.

Two conclusions can be drawn from this. First, different compensation filters are needed for different headphones. Theile (1986) and Møller et al. (1995a) formulated common design goals for headphones, but results from Lorho (2009) suggest that their rather theoretical approach does not yield the best perceptive results. This might be one reason for the variation among different headphone models. Second, the use of individual headphone compensation is essential if the transfer path from headphone to ear canal should be linearized. But note that a linear transmission not necessarily has to be perceptually best suited, if non-individual



Figure 2.9. HPTFs of 40 human subjects and 3 headphones taken from Møller et al. (1995b).

recordings are used.

Besides this effect, that is related to inter individual differences in binaural recordings, another aspect is of interest. Because HPTFs also depend on the way the headphone is placed on the head of a subject, considerable intra individual differences occur, if headphones are repositioned. Using a 3AFC test paradigm, Paquier and Koehl (2010) showed that intra individual differences were perceivable for any out of four headphones and three stimuli tested. However it has to be stressed, that binaural recordings have been made on an artificial head and that the headphones have been placed by the experimenter. Møller et al. (1995b) remarked, that intra individual variance is smaller, if the headphones are placed by subjects themselves. Unfortunately Paquier does only display averaged HPTFs. Therefore, the range of variability in his measurements is not known and it remains unclear if it is bigger or comparable to variances observed by others.

HPTFs from Schärer (2008) measured on the FABIAN HATS are depicted in Fig. 2.10. Ten measurements have been made for each headphone. The headphones have been repositioned by the experimenter between measurements, but variances are in good accordance with results from Møller et al. (1995b). For the *Stax SRS 2050 II* headphones, that are comparable to the *STAX SR Lambda professionals*, the leakage effect causes intra individual differences of about ± 3 dB below 200 Hz. For both headphones deviations are smaller than ± 1 dB up to 2 kHz. Above 2 kHz and apart from high deviations at frequencies where notches occur,



Figure 2.10. HPTFs measured on the left ear of the FABIAN HATS with Stax SRS 2050 II (left) and AKG K-1000 (right) headphones. Data from Schärer (2008).

deviations for the *STAX* reach ± 5 dB, whereas the *AKG* only shows ± 2 dB.

Measurements made on a HATS have been chosen to discuss intra individual variances, because influences of microphone positioning and head movements are excluded this way. Intra individual variances limit the precision of headphone compensation and can be perceived, as has been shown by Paquier and Koehl (2010).

2.5.3 Other Influences

ITDs and ILDs

The influence of spectral coloration induced by non-individual HRTFs and BRTFs was discussed before, and differences in pinna shapes were named as the main reason for this. In addition, and depending on the frequency, the head geometry influences ILDs. The differing energy on the contralateral ear, that was reported by subjects from Lindau et al. (2007) can possibly explained by this effect (see Chap. 2.2.4).

Moreover, head geometry influences ITDs which are the dominating cue for localization in the horizontal plane. If the binaural simulation allows for head movements, localization can become unstable, meaning that the source either moves according or contrary to the head movements of a listener (Lindau et al., 2010a). A relevant degradation of localization was not reported for static binaural synthesis (Wenzel et al., 1993; Møller et al., 1996b). However these studies used a relatively coarse grid of source positions, which might have been inadequate to evaluate small degradations induced by slightly differing ITDs.

Pressure division ratios

The FEC criterion, that has to be met if binaural signals are recorded at the blocked ear canal, has been introduced in Chap. 2.2.2. In average, four headphones meet the criterion with a tolerance of ± 2 dB. However individual measurements of the same headphones

showed bigger deviations of up to \pm 7 dB, and a compensation for this is not feasible in most cases. The best results were obtained with the extraaural *AKG K-1000* headphones, that only exhibited deviations of about \pm 3 dB (Møller et al., 1995b). On this account, violations of the FEC criterion may possibly result in potentially audible artifacts, if a binaural simulation is compared to listening in a natural environment.

Temporal course

Furthermore, Schärers subjects noticed differences in the transient or temporal behavior. It was assumed that this was caused by preringing from the linear-phase target function, as briefly discussed in Chap. 2.4. Regarding this, it has to be questioned whether the preringing induced by a linear-phase, or the group delay distortion induced by a minimum phase target response is perceptually best suited.

2.6 Chapter summary

In this chapter, binaural hearing and its simulation by means of binaural synthesis were introduced. It was shown that a plausible but not authentic simulation can be obtained, when non-individual recordings made with a HATS are used for auralization. The simulation still is distinguishable from a corresponding real sound field, mainly due to of spectral coloration induced by the non-individual recordings. An overview of the influences on the physical precision of binaural simulation in conjunction with emerging perceptual degradations is given in Tab. 2.4. Issues that were addressed in the current study are marked with an asterisk. The points (a)–(c) refer to the influence of binaural signals, (d)–(f) to that of the transducers involved in recording and reproducing such binaural signals, (g)–(i) to the precision the recordings can be made with and point (j) refers to the headphone compensation.

0	Drigin	Technical specifica- tion	Perceptual effect	Literature
(a) [*] Ir a:	nter individual vari- nce in HRTFs/BRTFs	Deviations from indi- vidual HRTFs/BRTFs up to ± 5 (20)** dB evoked by pinna, head and torso geometry (Fig. 2.8)	Coloration, lo- calization errors, degradation of spatial impres- sion	<i>CIPIC HRTF</i> <i>database</i> ; Møller et al. (1995c) – – – – – – – – Møller et al. (1996b); Schärer (2008)***
(b)* Ir a: (I ha ti	nter individual vari- ince in HPTFs In conjunction with leadphone compensa- ion)	Deviations from in- dividual HPTF up to ± 5 (20) dB evoked by pinna geometry. Distor- tion of linear transmis- sion (Fig. 2.9).	Coloration (gen- eral and ringing artifacts), local- ization errors	Møller et al. (1995b); Kulkarni and Colburn (2000) Møller et al. (1996a); Schärer (2008)
(c)* In au (I ho ti	ntra individual vari- ince in HPTFs In conjunction with leadphone compensa- ion)	Deviations of ± 3 (15) dB caused by repositioning of headphones. Distortion of linear transmission (Fig. 2.10).	Coloration, local- ization errors	Møller et al. (1995b); Schärer (2008) – – – – – – – – – Paquier and Koehl (2010)
(d) A	Acoustical loading of headphone (PDR)	Deviations of ± 2 (7) dB even for FEC headphones. Distortion of linear transmission.	No studies avail- able. Possibly coloration and lo- calization errors	Møller et al. (1995b) – – – – – – – – – none
(e)* H si	leadphone transmis- ion properties	Transmission limited at low frequencies due to electro acoustic trans- mission behavior	Coloration (poor bass) and pos- sibly missing tactile cues	Møller et al. (1995b); Schärer (2008) – – – – – – – – – – Schärer (2008)

continues on next page

	Origin	Technical specifica- tion	Perceptual effect	Literature
(f)	Loudspeaker directiv- ity	Missmatches from the source that is being simulated	Possibly col- oration and degradation of spatial impres- sion	Meyer (2008); Zot- ter (2009)
(g)	Recording/Repro- duction: Head move- ments and subject position	Deviations of ± 2 dB. Only affects HRTF and BRTF measurements	Coloration, if not carefully con- trolled .	Riederer (2004a) – – – – – – – – – – Hiekkanen et al. (2009)
(h)*	[*] Recording: Micro- phone position at earcanal	Deviations of ± 3 (10) dB	No studies avail- able. Possibly coloration	Riederer (1998, 2004a) – – – – – – – – – – none
(i)	Recording: Micro- phone response	Deviations of ± 2.5 dB above 12 kHz	Not reviewed. Possibly col- oration and lo- calization	see Chap. 3.1.3
(j)*	Phase response of headphone compensa- tion	Deviations from linear phase response (min- phase), or from tempo- ral behavior (lin-phase)	Preringing (lin- phase) or phase- ing (min-phase)	Norcross et al. (2006) – – – – – – – – – – Norcross et al. (2006); Schärer (2008)

Table 2.4 – continued from previous page

* Addressed in current study; ** mean value obtained from visual inspection (maximum value);

* * * Literature above dashed lines refers to technical specifications, literature below to corresponding perceptual effects.

Table 2.4. Influences on the quality of binaural simulation.

Chapter 3

Physical evaluation

Hence the evaluation of individual headphone compensation was the main aspect of this work, it was essential to have a tool for recording individual binaural signals at hand. It had to allow a fast and reliable in situ measurement and compensation of HPTFs, meaning both had to be done right before auralization. The development process for the measuring instrument started with the choice of the measurement position and further comprised an appropriate selection of measuring microphones and materials, that were needed for crafting. The crafting itself was followed by the development of the software, that controled the compensation \Rightarrow 3.1 Developing a measuring instrument. It ended with it's validation in terms of measuring and compensating individual transfer functions¹ \Rightarrow 3.2 Evaluation of headphone compensation.

3.1 Developing a measuring instrument

It was discussed in Chap. 2.3 that binaural recordings can be done at the open or blocked ear canal, and that none of the methods is superior to the other regarding the precision, that can be yield. However, if the measurements should be conducted in situ, custom ear moulds can not be used, as their production would take to much time. Furthermore, three reasons were in favor for measuring at the blocked, instead at the open ear canal: First, BRIR datasets were measured with the FABIAN HATS, that has microphones at the bottom of the cavum conchae at the point where the ear canal would begin, and the HPTFs should be measured at the same place. Second, Møller (1992) argued, that the least possible individual information is contained in the measurements at the blocked ear canal, and that this would therefore be the best choice if non-individual binaural recordings are used. Third, the level of the measuring signal at the eardrum is reduced by the blockage. That makes it possible to measure at higher levels and and better SNRs can be yield without having to average several measurements (Riederer, 1998).

For recording at the blocked ear canal, both, miniature and probe microphones could be used. However, the tip of probe microphones can not be considered rigid (Blauert et al., 1978), and medical tape would probably be needed to keep it in position. Thus miniature electret microphones were preferred.

¹A short description was given by Brinkmann and Lindau (2010)

Lastly, the material that holds the microphone in place and that is entered into the ear canal was chosen. A review of literature, that was done in Chap. 2.3, suggests two possiblities: compressible foam earplugs like the *E-A-R classic* or moldable silicon. The latter would have ensured a better repeatability of the microphone position but has the disadvantage, that the microphone would have to be pressed into the silicon for each measurement. This could have damaged the cables, that are usually soldered to the microphones backside, or silicon putty could have entered the microphone capsule opening, as Riederer (2004b) mentions. Since Riederer also reported problems using compressible foam earplugs, both materials did not seem to be the perfect choice. Instead, simply shaped silicon ear moulds, that should fit a large variety of ear canals were considered. If well designed, they should provide a blockage comparable to that of moldable silicon. If flush-casted into the ear moulds, the microphones would not have to be pressed into the earplug, as with moldable silicon and foam earplugs.

3.1.1 Anatomy of the human ear canal

For crafting the ear moulds, a good and precise knowledge of the geometry of the ear canal iss essential. However quantitative measures of the ear canal are barely reported in literature. In most cases, only measures of the head in general or of the pinna are given (Alexander, 1968; Burkhard and Sachs, 1975; Algazi et al., 2001).

An anatomic sketch of the outer ear is shown in Fig. 3.1 and a sketch of a cast of a human ear canal is depicted in Chan and Geisler (1990). The ear canal starts at the bottom of the



Figure 3.1. Anatomic sketch of the outer ear taken from Voogdt (2005). (Labels changed to english).

cavum conchae, is slightly directed upwards. It then passes the first and second bend, before it is terminated by the ear drum. It is entirely covered with skin, that first lies on a thin layer of cartilage and later directly on bone. Because of the cartilage at the beginning of the ear canal, it can adapt to an ear mould that is inserted into it to a limited amount. Average measures of the ear canal are given by Blauert (1997, p. 53), but these informations were not detailed enough for designing an ear mould. More data was provided by the hearing aid manufacture *PHONAK*¹. The measures given in Tab. 3.1 were extracted from 991 human ears, scanned with laser technology. For each measure, the mean, minimum and maximum value and the standard deviation is given. A legend of the measures is included in Fig. 3.2, which shows a 3D model of an outer ear up to the second bend.



Figure 3.2. 3D model of parts from a left outer ear. By courtesy of PHONAK.

	Mean	Min.	Max.	Std. Dev.
Major diameter aperture	14.50	9.60	21.94	1.82
Minor diameter aperture	7.75	3.71	13.99	1.39
1st leg	4.65	1.14	12.90	1.39
Major diameter 1st bend	11.79	6.81	19.88	1.74
Minor diameter 1st bend	7.29	3.34	14.59	1.40
2nd Leg	7.77	1.09	21.18	2.22
Major diameter 2nd bend	9.63	3.85	15.79	1.94
Minor diameter 2nd bend	6.27	2.11	10.39	1.36
Angle betw. 1st & 2nd leg	42.26	8.19	72.84	9.33

Table 3.1. Measures of the ear canal in mm and degree. By courtesy of PHONAK.

The diameters of the ear canal suggest that it's cross section is somehow oval, because the major diameter is always bigger than the minor. However, it can not be seen from the data, if and to what degree major and minor diameters are correlated. If they were not, other shapes would also be possible. But since Blauert (1997, p. 53) only reports circular and oval cross sections, a medium correlation can possibly be assumed. Towards the ear drum, the cross section then becomes more circular, which can be seen from the decreasing ratios of major and minor diameters. Not only the ratios of the diameters decrease towards the eardrum, but

¹See http://www.phonak.com (Last checked: May 2011)

also the diameters themselves. This means, that the ear canal commonly has a conical shape and narrows lengthwise towards the ear drum, which is also mentioned by Voogdt (2005, p. 30).

Another uncertainty is the distribution of the measures. Since many body dimensions, like height are normal distributed, and Algazi et al. (2001) reports measures of the pinna to be normal distributed as well, it seems reasonable to consider a normal distribution for measures of the ear canal, too. Assuming this, percentile values can be calculated using the z-score, given by

$$z_i = \frac{x_i - \bar{x}}{s}, \qquad (3.1)$$

where \bar{x} is the mean, *s* the standard deviation and z_i the z score belonging to a given area under the standardized normal distribution curve (Bortz, 2005, Chap. 2.5.1). This way, percentile values were calculated by solving Eq. 3.1 for x_i to obtain an overview of the amount of variability in the measures. The values shown in Tab. 3.2 cover 80 % of the

	10%	15%	25%	Mean	75%	85%	90%
Major diameter aperture	12.17	12.61	13.28	14.50	15.74	16.39	16.83
Minor diameter aperture	5.97	6.30	6.82	7.75	8.69	9.19	9.53
1st leg	2.87	3.20	3.72	4.65	5.59	6.10	6.43
Major diameter 1st bend	9.57	9.99	10.63	11.79	12.98	13.60	14.02
Minor diameter 1st bend	5.49	5.83	6.35	7.29	8.24	8.75	9.08
2nd Leg	4.92	5.46	6.28	7.77	9.28	10.08	10.61
Major diameter 2nd bend	7.15	7.62	8.33	9.63	10.95	11.64	12.11
Minor diameter 2nd bend	4.52	4.85	5.35	6.27	7.19	7.68	8.01
Angle betw. 1st & 2nd leg	30.31	32.55	36.01	42.26	48.60	51.96	54.20

Table 3.2. Mean and percentile values of ear canal measures in mm and degree.

population comprised by the 991 datasets. Within this fraction, the diameters end lengths vary for about 4 mm; the angle between first and second leg varies for 24°. Regarding this variation, it can be assumed that individual ear canals considerably deviate from the mean ear canal, if there is such. Due to this, not only one but three pairs of ear moulds were crafted in three different sizes, as described in Chap. 3.1.4 and 3.1.5. The selection of the microphones that were used for building the ear moulds, and the measurement and inversion of their frequency responses is described in Chap. 3.1.2 and 3.1.3. Minimum and maximum values were not considered for analyzing the variance in the measures, since they are possibly determined by outliers.

3.1.2 Microphones for binaural recordings

In the beginning of his chapter, it was argued that miniature microphones are best used for measuring at the blocked ear canal. A literature overview showed, that this narrows the possible candidates to only a few (see Tab. 3.3)¹. As the ear is an omnidirectional pressure

	Sennheiser KE 4-211-2	Knowles EA Series	Knowles FG Series	Panasonic WM 61 A
Frequency response	$\pm 3 dB$ between 0.02 and 20 kHz	$\pm 15 \mathrm{dB}$ between 0.1 and $10 \mathrm{kHz}$	flat between 0.1 and 10 kHz	$\pm 2 dB$ between 0.02 and 20 kHz
Sensitivity [mV/Pa @ 1kHz]	10	1.1	2.2	17.8
Equivalent noise level [dB(A) @ 1kHz]	27	28.5	30	n.s.
Capsule size [mm]	$4.2/4.75 \varnothing$	5.56/4.59/2.21	2.59/2.59	3.4/6 ø
Capsule geometry	cylindrical	cuboid	cylindrical	cylindrical
Directivity	omni	omni	omni	omni
Used by	1 – 2	3–5	6	7

1 Møller et al. (1995b); 2 Riederer (2004a); 3 Middlebrooks et al. (1989); 4 Chan and Geisler (1990) 5 Pralong and Carlile (1994); 6 LISTEN HRTF database (2002); 7 Rausch (2008)

 Table 3.3. Miniature microphones used for measuring at the ear canal. Specifications according to manufacturer.

transducer, only capsules with matching directivity were taken into account (Blauert, 1997, p. 54). The equivalent noise levels are comparable across the microphones, while the sensitivity varies, as it depends on the membrane diameter. However, it still was within an acceptable range, even for the *Knowles EA Series*. Regarding the diameters of the ear canal listed in Tab. 3.2, the most important criterion to be met by the microphone is its size, as diameters of 4.5 mm had to be considered for small ear canals. For this reason, the *Knowles FG-23329* was chosen, as it was the only microphone that met this criterion. It has about half the size of the widely used *Sennheiser KE 4-211* capsule. According to *Knowles*, the frequency response is flat up to 10 kHz. Above that, no specifications were given, but it was assumed that it can be equalized up to 20 kHz by means of digital filtering. The entire technical specifications can be found in Appendix A. Similar to the *KE 4-211*, the *FG-23329* capsule is enclosed by a metal housing, that protects the microphone membrane. This way, it should be robust against mechanical stress and should not sustain damage from being pressed into the ear canal.

¹Wightman and Kistler (1989) used an *ETYMOTIC RESEARCH* miniature microphone, without specifying the exact type. Thus it is not listed in Tab. 3.3.

Binaural microphones for consumer use, like the *Sound Professionals MS-TFB-2* or the *Bruel & Kjær 4101* were not considered, because it was assumed that they (a) provide an insufficient blockage, (b) can not be inserted into the ear canal and (c) don't allow for establishing a defined measuring position at the ear canal.

3.1.3 Microphone measurement and inversion

The inverse frequency responses of the microphones used for recording HPTFs and BRTFs are included in the compensation filter, that has been introduced in Chap. 2.2.2 and 2.4. It was mentioned, that their inversion is easier than that of the HPTF, but for completeness it is described in the following.

The frequency responses of the microphones were measured in the anechoic chamber of the Institute for Technical Acoustics and Fluid Mechanics of the Technical University Berlin. It has a lower cut-off frequency of 63 Hz, below, the free field sound transmission is disturbed by room modes and measurements are not valid.

The free field frequency responses of six *Knowles FG-23329* miniature electret condenser microphones have been determined using a *Monkey Forest* measuring system ¹ and the substitution method, similar to the description given by Müller (n.s. b). This way, the frequency responses were determined with respect to a reference microphone with known properties. Therefore the frequency response of the reference microphone was measured first. While the reference microphone remained in place, the small FG-23329 capsules were attached to it and fixed with adhesive tape, to assure that the measuring position was maintained. Then, their frequency responses were measured as well. A *Brüel & Kjær*¹/4" free field microphone type 4135 with tolerable deviations from a flat frequency response between 0.02 and 20 kHz was used as reference in conjunction with a *Brüel & Kjær* measuring amplifier type 2610 (see Appendix A). Sine sweeps of order 15 with bass emphasis were used as excitation signals and circular deconvolution was employed to obtain the impulse response from the measured signals (Müller and Massarani, 2001; Müller, n.s. a). A *Genelec 8030a* two way studio monitor was used to play back the excitation signals, and the microphones were placed 4.54 m in front of it, on-axis with the tweeter (see Appendix A).

Reliable measurements can only be made in the far field of a loudspeaker, where its frequency response only depends on azimuth and elevation angels between the loudspeaker and microphone. The two relevant far field criteria are given by Möser (2009, Chap. 3.5.4). The third criteria can be disregarded, when pressure transducers are used.

(a) The sound pressure that is radiated from the loudspeaker decreases with distance. This

¹See http://www.four-audio.com/de/produkte/monkey-forest.html (Last checked: May 2011)

decrease has to be identically for sounds radiated from different loudspeaker units and reflections from the loudspeaker housing, and therefore the distance to the microphone has to be large compared to the source dimension:

$$r_{meas} \gg h$$
 .

The largest source dimension was the diagonal of the loudspeaker (0.35 m). In this case the measuring distance was more then the tenfold of the source dimension.

(b) The loudspeaker only radiates a directivity pattern, that is independent from the source distance, if phase differences induced by different travel paths from its units to the microphone are small. This is given for:

$$r_{meas} \gg rac{h^2}{\lambda}$$
 .

If solved for *f* ($\lambda = c/f$), an upper cut-off frequency of 12.6 kHz is given, below which the measurements are valid.

Since the measurements should be conducted up to 20 kHz, the second criterion is violated. However this can be disregarded if reference and measured microphones are located at exactly the same place.

The magnitude responses of the six FG-23329 microphones are shown in Fig. 3.3(a), for completeness, the raw measurements are given as well. They were obtained by dividing the complex spectra of the FG-23329's and the reference microphone.

In accordance to the specifications given by *Knowles*, the responses are almost constant up to 10 kHz, followed by a slope. At 20 kHz, the responses are 5–9 dB below the previous level, which are acceptable and correctable deviations. The raw measurement are unsmooth due to SNR problems in the low, and to reflection in the high frequency range. Thus, the following steps were applied in post-processing to both, the reference and FG-23329 measurements¹:

- (a) The acoustic delay, induced by the distance to the loudspeaker was eliminated by a circular shift of the impulse response. The point, where the impulse starts, was determined with respect to the maximum value. After visual inspection, values of -35 to -24 dB have been chosen.
- (b) The impulse responses were shortened to 128 samples, by applying a right sided tukey window, using the tukeywin function of Matlab©. The ratio of taper to constant sections

¹The post-processing was done in *Matlab*©; the source code is appended in D.



Figure 3.3. Knowles FG-23329 frequency responses. (a) Raw and processed magnitude response of 6 microphones, with 15 dB offset for clarity; (b) Phase responses in black and minimum phase responses in red for pairs of microphones; (c) Phase differences for pairs of microphones; (d) Magnitude responses and time signals of inverted microphones responses.

was set to 0.7.

- (c) Fractional octave smoothing with a width of a sixth of an octave was used to eliminate reflections at high frequencies that remained after windowing. This was done based on the smoothnew3 routine by Welti¹.
- (d) The phase response was set to minimum phase using the Hilbert transform.
- (e) The frequency response was inverted with a limited dynamic of 100 dB. This means that every value that lied more than 100 dB below the maximum, was clipped before inversion.
- (f) The responses were corrected for the reference by dividing the complex spectra.
- (g) The gain was adjusted, for the least sensitive microphone to have a level of 0 dB at 1 kHz.
- (h) The magnitude responses were set to a constant below 350 Hz, because they suffered from the windowing applied in a previous step. Pressure transducers can in general be assumed to have a constant frequency response according to their electro acoustic properties (Möser, 2009, Chap. 11.1).

The magnitude responses resulting from the post-processing, excluding step (e), are shown in the bottom of Fig. 3.3(a). The sensitivities of the microphones differed for about ± 2 dB; two microphones had a relatively low sensitivity compared to the others and a slope that started later but was steeper. In general, differences between them were in the range specified by the manufacturer.

Problems may arise when using the minimal and discarding the original phase of the microphones. This was done in step (d), because minimum phase systems are easier to handle when they should be inverted. However this way, phase differences between microphone pairs was not corrected. Absolute phases are shown in Fig. 3.3(b) and phase differences in Fig. 3.3(c) for pairs of microphones that were used for a left and right ear mould, respectively. They were obtained following the procedure described above without applying step (d). Phase differences are only evaluated by the auditory system below 1.5 kHz. In this range, two of the three microphone pairs exceededed the IPD threshold given in Tab. 2.1. This error could be avoided in future works by designing linear phase microphone filters as discussed in Chap. 2.4.1.

¹See http://www.mathworks.com/matlabcentral/fileexchange/26771-figutils/content/smoothnew3.m (Last checked: July 2011).

The microphone filters are depicted in Fig. 3.3(d) and since the frequency responses are minimum phase, they simply were obtained by calculating the reciprocal of the processed transfer functions.

Besides the *Knowles* microphones, that will be casted into the ear moulds, the *DPA 4060* microphones mounted in the FABIAN HATS were also measured and inverted, following the procedure described above. Results are presented in Fig. 3.4 and are comparable to that of the *Knowles* microphones. The frequency responses of the DPA 4060's are constant up to 1 kHz. Then, a soft boost with a peak at 9 kHz and a gain of 5 dB appears. The phase differences slightly exceeds the threshold given before, but it is questionable, if this is audible with any other signals than pure tones.

Power supply for Knowles FG 23329 microphones

The *Knowles* microphones have to be supplied with a voltage from 0.9–1.6 V. Since the wiring of the microphone capsule does not allow the supply with phantom power, even if the voltage is reduced by a voltage divider, a custom power supply was build. It is powered by a AA battery, and was designed to work with *Knowles* and *Sennheiser KE 4-211-2* microphones. The magnitude response of the power supply was measured with *Monkey Forest* and showed no noteworthy deviations from a constant magnitude. The circuit diagram can be found in Appendix A.

Repeatability of microphone measurements

The frequency responses of the *DPA* microphones, that are build into the FABIAN HATS were already determine in a previous study by Lindau (2006), which makes is possible to discuss the two measurements regarding errors induced by the measuring procedure. It has to be stressed however, that this is not a systematically examination of possible error sources, and that there is not enough data available to draw reliable conclusions from it. Rather, it is providing a rough estimate of errors that might occur.

Lindau measured the frequency responses using the substitution method, *Monkey Forest* and a *Büel & Kjær* microphone amplifier of type 2610. The magnitude responses of the *DPA 4060* are shown in the bottom and difference plots are given in the top of Fig. 3.5. In general, the measurements are in good agreement to each other. For the right microphone, errors are below ± 1 dB within the whole frequency range. For the left microphone, they are in the same range up to 12 kHz, above, deviations of +2.5 dB can be observed. The error seems to be somehow systematically, since the deviations are negative between 1 kHz and 3 kHz, and positive above 10 kHz for both microphones. A possible source for this error is the positioning of the *DPA* microphones with respect to the reference microphone.



Figure 3.4. DPA 4060 frequency responses. (a) Raw and processed magnitude response of 2 microphones, with 10 dB offset for clarity; (b) Phase responses in black and minimum phase responses in red for pairs of microphones (left) andhase differences for pairs of microphones (right); (c) Magnitude responses and time signals of inverted microphones.



Figure 3.5. Free field frequency responses of DPA 4060 microphones from left and right ear of FABIAN measured in 2006 (red) and in 2011(black, bottom); Difference between successive measurements calculated on a dB basis (top).

As well as the variability in HPTFs described in Chap. 2.3, the variability in the magnitude responses of the transducers used for measuring binaural signals limits the precision of the binaural synthesis.

3.1.4 Prototyping

Before the measuring instrument was crafted, a first prototype of the ear mould and later a negative mould was designed using the 3D modeling software *Blender*¹. Strictly speaking, *Blender* is not intended for prototyping and does not provide much tools for measuring the sizes of shapes and objects. But most operations like specifying the size of simple geometric objects, or extruding, moving and rotating objects can be done by typing in exact values for the operation via keyboard. Further *Blender* does not provide a unit for the objects that are designed, but it was found out that one *Blender* unit equals one millimeter. Despite these shortcomings, *Blender* was used for prototyping, because it is non-proprietary.

First prototype

The first prototype of a left ear mould with flush-cast microphone is depicted in Fig. 3.6. The major and minor diameters of the front and back side of the prototype equate to the mean values of first and second bend diameters, as given in Tab. 3.2. These measures were preferred to the aperture diameters, because the length of the first leg was considered to short to fit the ear mould including the microphone capsule. This way, the ear moulds are

¹See http://www.blender.org/ (Last checked: May 2011)



Figure 3.6. Prototype of left ear mould with flush-cast microphone.

a bit smaller than necessary and hopefully provide a good fit in spite of the first bend. As mentioned earlier, the diameters indicate that the ear canal has a somehow oval cross section, but the exact shape can not be drawn from them. The cross-sectional shape of the ear moulds, that can be seen in the left side of Fig. 3.6 was recreated to custom ear moulds found in the archive of the Audio Communication Group (see Appendix C). Based on them, the side of the ear moulds, that points to the frontal direction when inserted into the ear canal (left side of left ear mould in Fig. 3.6) was designed flatter than the opposite side.

A second uncertainty regarding the shape of the ear moulds, was the position of the back side with respect the the front side. It can not be seen from the data provided by *Phonak* whether or not the intersections of major and minor diameters of either sides of the ear moulds are horizontally and vertically aligned or not. Thus the orientation of the sides was again recreated to the custom ear moulds found in the archive, which showed a slight vertical but no horizontal displacement. This can be seen from the middle and right prototype in Fig. 3.6.

The shape of the ear mould, as described above, was obtained starting with a circle and scaling it to match the wanted major and minor diameters. Its shape was then manipulated to recreate the cross section of the ear moulds from the archive. In the next step the face was extruded to a three dimensional object with the length of the second leg, and in a final step the back side was scaled and vertically displaced.

Negative mould

It was discussed in Chap. 3.1.1, that the cartilaginous part of the ear canal can adapt to an ear mould to a certain amount, but that due the variation in the measures of the ear canal, ear moulds should be crafted in multiple sizes to provide a good fit to various ear canals. Therefore, a negative mould with cavities of five different sizes has been designed in *Blender* which is shown on the left side of Fig. 3.7. The measures that were used for the negative



Figure 3.7. Negative mould. Mesh Grid (left) and 3D plot (right).

mould are given in Tab. 3.4. It was considered that a length of at least 10 mm is needed to anchor the microphone in the ear mould, and thus it was set to 10 mm whenever the length of the second leg was shorter than this.

Size (Percentile value)	xs (15%)	s (25%)	m (50%)	1 (75%)	xl (90%)
Major diameter 1st bend	9.99	10.63	11.79	12.98	14.02
Minor diameter 1st bend	5.83	6.35	7.29	8.24	9.08
Length	10	10	10	10	10.61
Major diameter 2nd bend	7.62	8.33	9.63	10.95	12.11
Minor diameter 2nd bend	4.85	5.35	6.27	7.19	8.01

Table 3.4. Measures used for negative mould (uncolored) and final measuring instruments (colored).

The negative mould was then 3D-printed with a *Contex Designmate Cx* at the 3D-Laboratory of the Institute for Mathematics at the Technical University Berlin¹. With this method, thin layers of plaster are adhered with water based binder. The precision is given by the thickness of the layers, which was 0.0875–0.1 mm. After the printing, the model is infiltrated with a two component mixture consisting of epoxy resin and hardener to make it resistant to mechanical stress. The negative mould is depicted on the right side of Fig. 3.7.

3.1.5 Crafting

Before the ear moulds were crafted, three things had to be discussed: What material can be used for casting, how many different sizes are needed to cover a vast variety of ear canals and how can the microphone capsules be anchored in the ear moulds.

¹See http://www.tu-berlin.de/3dlabor/ (Last checked: June 2011)

The material used for casting should be compressible to a certain degree, so that the ear moulds can adapt to the form of the ear canal (and vice versa). Furthermore, it had to be self curing at room temperature. Materials, that harden by heat supply, pressurization or in a water bath can not be used, because this would have possibly damaged the microphone capsules. An overview of materials and production methods for ear moulds is given by Voogdt (2005) and only addition-curing RTV-silicone and light-curing materials satisfy these criteria. Light-curing materials, however, have been excluded from further considerations because special devices are needed for curing.

For crafting ear moulds, commonly two-component addition-curing silicones are used. They are water repellent, mostly skin compatible and resistant to chemical substances, elastic and consistent in form, and shrink less than 0.1 % when curing (Voogdt, 2005, pp. 96). However processing them is difficult, because the specified mixture ratio has to be met exactly and all tools have to be thoroughly cleaned. Therefore, one-component silicone that is usually used as sealing material in sanitary areas was taken for the first prototypical ear moulds. It has similar properties as two-component silicone, but needs much longer for curing.

Before crafting ear plugs with flush-cast microphones, solid silicone ear plugs in all five sizes (xs, s, m, l, xl) were made, and after informal tests it was decided that the smallest, and largest size could be discarded. The crafting process of the prototypical ear moulds is illustrated in Fig. 3.8. The microphone cables were bend 180° directly behind the microphone capsules and held in place with little pieces of shrink tube with a diameter of 4.5 mm. The shrink tube was carefully heated without damaging the microphone. Anchors were made from two small hollow metal balls with a diameter of 2.4 mm attached to either side of a fishing line (0.5 mm diameter). The metal balls were carefully bend open and then pressed onto knots at the end of the fishing line. The anchor was then placed at the bend of the cable and kept in place by the shrink tube. For the size s ear mould, an anchor consisting of only one metal ball was constructed to fit the negative mould, as shown on the right side of Fig. 3.8(a).

Before casting the ear moulds, the negative mould was once lacquered with *Silon* from *Dreve*¹ and little wholes were drilled into the ground to support curing. In addition, a thin layer of petrolatum was applied, to ensure the silicone mould could be easily separated from the negative form after it had cured. Then, the microphones were fixed on aluminum round-bars with adhesive tape and brought in the right position. Finally the silicone was carefully injected into the negative mould. Surplus silicone was removed, and the surface of the ear mould was smoothed using a small metal plate, that was also treated with petrolatum (see Fig. 3.8(b)).

¹See http://www.dreve.de/dreve_neu/otoplastik_gb/service_oto_gb.htm (Last checked: June 2011).



Figure 3.8. Crafting of the measurement instrument. (a) Microphone preparation; (b) ear mould casting; Ear moulds: (c) small, (d) medium, (e) large.

After seven days, the ear moulds could be separated from the negative mould. In a last step, they were cleaned with disinfectant and surplus material was removed with a sharp knife. The finished ear moulds are depicted in Fig. 3.8(c)–3.8(e).

3.1.6 Physical evaluation of the measuring instrument

As discussed in Chap. 2.3, binaural measurements are very sensitive towards microphone positioning and displacements of 1–2 mm may induce an error of ± 10 dB. Therefore, a physical evaluation of the measuring instrument and a quantification of the measurement error induced by displacement of the microphone capsule was desirable. This could have been assessed by measuring multiple HRTFs or HPTFs of one subject and reinserting the measuring instrument into the ear canal between successive measurement. However this way, either head movements or intra individual variance in HPTFs would have induced additional errors. For these reasons, the evaluation was carried out using an artificial ear with ear canal from *Dreve*.

3.1. DEVELOPING A MEASURING INSTRUMENT

To obtain an overview of different methods, measurements were conducted using the measuring instrument as described before, as well as two compressible foam ear plugs, where the microphones were placed in slits that were cut into the foam. Ten measurements have been conducted for each method and the ear plugs have been removed from the ear canal and reinserted between successive measurements. The artificial ear was fixed on the ground with double sided adhesive tape and sine sweeps were played back through a loudspeaker placed 1.5 m above the ground. Measurements were conducted with *Monkey Forest* in an office room. Unwanted reflections have been eliminated in post processing by means of windowing and smoothing, and the frequency responses were corrected for the microphone by applying the filters described in Chap. 3.1.3.



Figure 3.9. Measurements conducted on an artificial ear with ear canal. (a) Microphone placements with different materials. (b) Corresponding magnitude responses for ten measurements.

Exemplary positions of the microphone capsule in the artificial ear are depicted in Fig. 3.9(a). On the left side, the capsule was inserted into an *E-A-R Classic* and in the middle into an *Uvex com4-fit* foam ear plug. The measuring instrument used throughout this study is shown on the right side. In contrast to the silicon ear plug (size s), the foam ear plugs could not be

inserted entirely into the ear canal, because they were to big. For this reason it was hard to establish a well defined measuring position with the foam ear plugs.

The corresponding magnitude responses are shown in Fig. 3.9(b) for frequencies above 2 kHz. Below, measurements were disturbed by reflections from the room, however this can be disregarded, because variations in binaural recordings due to microphone positioning are generally small at frequencies below 2 kHz (see Chap. 2.3). Since the measurements were not calibrated, absolute sound pressure levels are not given, but the levels agree across the measurements. All magnitude responses exhibit peaks and notches that originate from pinna resonances. They are most pronounced for the silicon ear plug, which indicates a good blockage of the ear canal (Riederer, 2004b). The smallest variabilities within the measurements are seen for the silicon ear plug. They are negligible below 8 kHz, and above, maximum deviations of ± 2 dB occur, whereas deviations of up to ± 15 dB can be observed with the foam ear plugs.

In summary, the silicon ear plugs ensure a higher reliability than commonly used foam ear plugs and measurements with them are easier to conduct. The measuring instrument has been named **p**recisely **re**peatable a**c**quisition of **i**ndividual binaural tran**s**f**e**r functions (PRECISE).

3.2 Evaluation of headphone compensation

Before the influence of individual headphone compensation was assessed in a listening test, a physical evaluation was carried out. Therefore individual HPTFs were measured on 25 subjects. The HPTFs were then compensated as described in Chap. 2.4 and their deviations from the target bandpass were examined by means of an auditory filter bank. This way, the effect of different headphone compensations on the binaural transmission chain could be estimated. In addition, individual HPTFs were needed before the perceptive evaluation, to obtain a generic headphone compensation filter. Further, this also provided a good opportunity for testing wether or not the PRECISE ear moulds fit a large variety of ear canals.

3.2.1 In situ HPTF measurement

Individual HPTFs were measured on 25 subjects (21 male, 4 female, avg. age 31 years) with circumaural *STAX SRS 2050 II* headphones. For reasons of feasibility, no attention was paid on drawing a representative sample regarding gender or age. The *STAX* headphones were chosen, because they were subjected to perceptual tests by Schärer (2008) and are known to meet the FEC criterion better than other circumaural headphones (Møller et al., 1995b). Ten

transfer functions were measured per subject and the headphones were repositioned between successive measurements by the subjects. Measurements were conducted with the FABIAN software using sine sweeps of order 16 (Lindau, 2010; Müller and Massarani, 2001).

The PRECISE ear moulds were inserted by the subjects with help of a little plastic stick that was used to to press them into the ear canal. Afterwards, the position was visually inspected by the experimenter and corrected by the subject if necessary (see Fig. 3.10).



Figure 3.10. PRECISE ear moulds in human ear canals. (a) small, (b) medium, (c) large

The small ear moulds were used in 21 cases, whereas the medium were used three times and the large only once. No subject complained about the small ear moulds being to large or the other way around. Subjects 10, 12 and 24 reported the left ear mould to be slightly lose, while the right one had a good fit. All three used small ear moulds, because the medium size did not fit their ear canals. For subject 10, the ear mould had to be repositioned during the series of measurements. However, HPTFs of these subjects showed no abnormalities compared to others, which might be seen as another proof for their reliability.

Since the blockage of the ear canal decreases the level of the measuring signal at the ear drum, higher measuring levels could be used, and a sufficient SNR was obtained without averaging multiple measurements. This way, measuring ten individual HPTFs took between five and ten minutes, including the insertion of the ear moulds. Before a measurement series was finished and the subject removed the ear moulds, the measured HPTFs were displayed with a prepared *Matlab*© routine and visually inspected. In case of unusual results, the measurement series could have been repeated. This however had not to be done for neither of the 25 subjects. After each series, the ear moulds were cleaned with disinfectant.

Intra individual variance

The intra individual variance in HPTFs limits the precision of the headphone compensation, as briefly discussed in Chap. 2.5.2. It is caused by repositioning of the headphone, and its amount depends on the headphone that is used, as well as on the subject. The effect of the latter will will be discussed in the following.



Figure 3.11. 10 HPTFs measured on two subjects (grey) and average HPTF (black) for left and right ear. (a) good repeatability, (b) poor repeatability.

HPTFs of two exemplary subjects, that cover the range of intra individual variances, are shown as grey lines in Fig. 3.11. The average HPTF is depicted black ¹. Below 50 Hz, no estimate of the intra individual variance can be given, because the SNR is insufficient for the non-averaged HPTFs. However, this does not affect the headphone compensation, because (a) the average HPTFs are used for the calculation of the compensation filter and the averaging eliminates most of the low frequency noise, and (b) the compensation range is restricted by the target bandpass with a lower –6 dB cut-off frequency of 50 Hz. HPTF measurements of subject 17 are depicted in Fig. 3.11(a). In this case, deviations are smaller than ± 1 dB below, and ± 2 dB above 6 kHz. Bigger deviations only occur with notches at 9 kHz and 15 kHz, that

¹HPTFs of 95 subjects have been measured during this work. Graphs of them can be found on the attached CD (see Appendix D).

are caused by anti resonances of the pinna. The center frequencies of the notches are almost perfectly reproduced and only the gains differ. The HPTFs of subject 20 that are depicted in Fig. 3.11(b), exhibit more intra individual variation. It already reaches ± 2 dB below 200 Hz, probably originating from leakage (see Chap.2.5.2, and Dillon, 1977). Between 200 Hz and 2 kHz, deviations are comparable to subject 17, but above they exceed ± 10 dB, caused by shifting center frequencies of notches at approximately 9 kHz, 15 kHz and 19 kHz. The observed variances are in good accordance to other studies and are likely to be audible, at least for cases with poor repeatability (Møller et al., 1995b; Kulkarni and Colburn, 2000; Paquier and Koehl, 2010).

Inter individual variance

The inter individual variance is of interest, because it limits the precision that can be yield with non-individual or generic headphone compensation and was as well discussed in Chap. 2.5.2. Mean HPTFs of 25 subjects are depicted in Fig. 3.12(a). They have been normalized to 0 dB at 300 Hz, to emphasize the common structure with peaks at about 1.8 kHz, 5.5 kHz and 10 kHz, and notches at 8 kHz and 15 kHz, that can be observed despite the variance. Before the normalization, the mean HPTFs exhibited deviations of ± 1 dB at 300 Hz, which is well comparable to results from Møller et al. (1995b).

In Fig. 3.12(b), the variance is illustrated by the 12.5%–87.5% percentile range with respect to the overall mean HPTF. It was calculated on the basis of all 250 HPTFs measured on left ears. Four characteristic frequency ranges can be identified from that. Below 200 Hz, differences of –6 dB to +3 dB can be observed, which can primarily be assigned to leakage effects. Since leakage causes a peak at about 100 Hz and and a steeper slope, which can be seen in Fig. 3.12(a), the deviations from the mean HPTF are asymmetrical. Negative deviations predominate below 50 Hz, and positive between 50 Hz and 200 Hz. Up to 2 kHz, deviations are smaller than ± 1 dB. Above 2 kHz, and up to 5 kHz, they quickly increase to ± 3 dB. Above 5 kHz, the region of narrow pinna notches begins. Hence, deviations again become asymmetrical and reach –11 dB and +5 dB, respectively. Results for left and right ears are in general comparable, however stronger leakage occurred for the right ears (Brinkmann and Lindau, 2010).

3.2.2 Headphone compensation

The theory of headphone compensation for binaural synthesis was discussed in Chap. 2.4. Here, only practical issues regarding the parameters that were used in comparison to Schärer (2008) and the in situ inspection of the compensation results will be mentioned.



Figure 3.12. Inter individual variance in HPTFs. (a) mean HPTFs of 25 subjects (grey) and overall mean calculated on a dB basis (black) for left ears, (b) 12.5%–87.5% percentile range of left ears with respect to overall mean HPTF.

The headphone compensation and the inspection of the results were realized in *Matlab*©, based on routines that were kindly provided by Schärer, Lindau, Schulz and Rotter¹. The source code is maintained in a subversion directory² and can be found on the CD attached to this work (see Appendix D). In the following, the processing steps as well as the parameter values and routines that were used are described. Six different filters , two individual, two generic and two non-individual, were calculated for the physical evaluation. Each filter was calculated using a minimum and a linear phase target function from regularization. The main function that organizes the compensation is calc_hp_filter.m, which applies the following steps:

- (a) Compensate microphones used for measuring HPTFs (see. Chap. 3.1.3).
- (b) Shorten HPTFs to 2¹¹ samples and apply right sided cosine window with a length of 2⁹ samples. Complex averaging in frequency domain, separately for left and right channel

¹Rotter (2010); Schultz (2011)

²https://srv2.ak.tu-berlin.de/alindau/postpro/7 hp filter calculation/

(shorten_multiple_wavs, average_multiple_wavs).

- (c) Normalize average HPTFs at 300 Hz to -6 dB to avoid clipping of the compensation filters, which are later saved as wav-files. Slightly differing gains are applied for left and right channel induced by differing sensitivities of left and right headphone unit.
- (d) Fractional octave smoothing of averaged HPTFs based on the smoothnew3 routine by Welti¹ (Optional, fract_oct_smooth).
- (e) Calculate inversion filter based on average HPTFs. Separately for left and right channel (see. Chap. 2.4).
 - i Calculate target bandpass and normalize to -6 dB to match level of HPTFs (see step (c)).
 - ii Calculate amplitude regularization function. Separately for left and right channel
 (shelve2_SCF).
 - iii Calculate headphone compensation filter according to Norcross et al. (2006), with desired phase response (get_MinPhaseTarget_FFDInverseFilter, make_phase).
- (f) Correct gain difference between left and right channel of average HPTFs applied in step (c).
- (g) Compensate microphone used for measuring HRTFs/BRTFs (see. Chap. 3.1.3).
- (h) In case the compensation filters (time domain) contain values greater than one, all steps are repeated and the normalization level applied in step (c) is decreased by a pre-calculated value.
- (i) Save compensation filters as 32 bit wav-files.
- (j) Visual inspection of compensation results; adjustment of beta values and recalculation of compensation filters, if necessary.

The different processing steps are mainly independent from each other, so that different target or regularization functions or even other inversion approaches could easily be implemented, if desired. The calculation of the compensation filter is comparable to Schärer (2008), however it differs in three respects: First, Schärer did not apply steps (a) and (g), because the same microphones were used for measuring BRTFs and HPTFs. As mentioned in Chap. 2.4, this is incorrect in a mathematical sense, but does not seem to have much

¹See http://www.mathworks.com/matlabcentral/fileexchange/26771-figutils/content/smoothnew3.m (Last checked: July 2011).

influence on the compensation results, as can be seen from control measurements in Schärer and Lindau (2009, Fig. 11). However, it influences the effect of the beta weights on the compensation: The beta weights limit the amount of work, i.e. the amount of gain, that is applied to compensate the parts of the HPTFs whose level is below that of the target function (see Chap. 2.4, and Norcross et al. (2006)). The magnitude response of the DPA microphones, that were used during Schärers study, show a boost of up to 5 dB between 3 kHz and 19 kHz (see Fig. 3.4). If this is not corrected before the filter calculation, notches within this frequency range are boosted, and thus less affected by the beta weights. Schärer used beta weights of 0.4, but using the same values caused to much regularization for the HPTFs measured in this study. Therefore the beta values have been adjusted, to obtain compensation results, comparable to those of Schärer, which is the second difference. Third, Schärer designed a linear phase target function with a length of 2049 samples that is symmetric around sample 1025, whereas the last sample was cut from the target function in this work. This way, the target function does not exhibit a linear phase in a strict system theoretical approach, and will be referred to as unconstrained phase. Linear phase filters can of course also be designed for an even number of taps, however this has not been done, because the unconstrained phase target bandpass only showed negligible group delay distortions of ± 1.25 samples between 0 Hz and 20 kHz.

Based on ten HPTFs measured on FABIAN a non-individual, and based on the 250 individual HPTFs a generic compensation filter has been calculated with the described procedure. Both are depicted in Fig. 3.13. For the calculation off all compensation filters, a low-shelve with a mid-gain frequency of 4 kHz and a gain of -15 dB has been deployed. The beta weights, that were used are shown in Tab. 3.5. In the individual case, the beta weights have been adjusted to avoid high overshoots in the compensated HPTFs. To achieve this, different values had to be applied, depending on the individual measurements.

Beat weights	Individual mean (min/max)	Generic	Non-individual
Left ear	0.09 (0.02/0.2)	0.2	0.2
Right ear	0.11 (0.05/0.3)	0.2	0.2

Table 3.5. Individual, generic and non-individual beta weights.

3.2.3 Compensation results

Exemplary individual compensation results for subject 17 are depicted in Fig. 3.14, graphs for all 25 subjects are given in the *Plots* section of Appendix D. Starting with a comparison



Figure 3.13. (a) Non-individual compensation filter. HPTFs measured on FABIAN (grey) and mean HPTF (black), compensation filter (red) target bandpass (dashed). (b) Generic compensation filter. Mean HPTFs of 25 subjects (grey) and overall mean (black), compensation filter (red) target bandpass (dashed).

of unconstrained and minimum phase compensation, it can be seen from Fig. 3.14 (a) and (b), that the unconstrained filter leads to imprecise results below 100 Hz. Schärer (2008) mentioned that this problem could be solved by using higher filter orders, but that it would increase the delay induced by the linear phase approach. The group delay is given in 3.14 (c). Above 200 Hz both approaches lead to a constant group delay, which is 23.2 ms (1024 samples) for the unconstrained phase and 2.3 ms (100 samples) for the minimum phase compensation. At frequencies below 200 Hz, both methods exhibit monaural group delay distortions, that exceed the 0.5 ms threshold given by Blauert and Laws (1978)¹. While the unconstrained phase compensation leads to nearly symmetric group delay distortions of ± 2 ms above 40 Hz, the minimum phase compensation exhibits distortions of +15 ms in the same range. The phase differences between left and right channel are shown in Fig. 3.14 (d). The results are comparable among the two methods, and between 40 Hz and 500 Hz, deviations slightly exceed the threshold of 4° given by Mills (1958) (see Chap. 2.1). Above 500 Hz, noise in the phase response, complicates a discussion. The impulse responses

¹Citet after Schärer (2008, p. 81)



Figure 3.14. Individual headphone compensation of subject 17, left ear. Unconstrained (left column) and minimum phase (right column). (a) Individual HPTFs and compensation filter; (b) compensated HPTFs; (c) group delay of left and right channel; (d) phase difference between left and right channel; (e) impulse response of compensated HPTF. ((c) - (e) are taken from one exemplary HPTF).



Figure 3.15. Generic headphone compensation of subject 17, left ear. Unconstrained (left column) and minimum phase (right column). (a) Individual HPTFs and generic compensation filter; (b) compensated HPTFs; (c) group delay of left and right channel; (d) phase difference between left and right channel; (e) impulse response of compensated HPTF. ((c) - (e) are taken from one exemplary HPTF).



Figure 3.16. Non-individual headphone compensation of subject 17, left ear. Unconstrained (left column) and minimum phase (right column). (a) Individual HPTFs and non-individual compensation filter; (b) compensated HPTFs; (c) group delay of left and right channel; (d) phase difference between left and right channel; (e) impulse response of compensated HPTF. ((c) - (e) are taken from one exemplary HPTF).

are shown in Fig. 3.14 (e). The unconstrained phase impulse response exhibits a bigger delay and more prominent pre-ringing. The minimum phase impulse response has a delay of 100 samples that has been manually introduced to shift pre-ringing artifacts to the beginning of the impulse response.

Generic and non-individual compensation results of the same subject are given in Fig. 3.15 and 3.16. Aside from obvious deviation of the compensated HPTFs from the target bandpass, that will be discussed in the next section, bigger deviations in the group delay and phase responses can also be observed, especially with the non-individual compensation approach.

3.2.4 Auditory modeling of compensation results

So far, thresholds, above which deviations become audible were only introduced for monaural group delay distortions and binaural phase differences. In this section, the auditory effect of deviations of the magnitude responses of the compensated HPTFs from the target band pass will be considered. This was done using an auditory filter bank that models the behavior of the human auditory system. As shown in Fig. 3.17, it is constructed from 40 overlapping equivalent rectangular bandwidth (ERB) filters according to Moore (1995). It was realized using MakeERBFilters and ERBFilterBank from the Auditory Toolbox for Matlab© (Slayney, 1998). The auditory modeling was done as described by Schärer (2008, pp. 62,



Figure 3.17. Auditory filterbank consisting of 40 ERB filters.

Eq. 4.6): For each subject, the target band pass and ten compensated HPTFs were filtered by each band of the filter bank and the difference between compensation result and target bandpass was calculated on a dB basis.

For all six compensation approaches, deviations from the target functions are depicted in Fig. 3.18. As a rule of thumb, it can be assumed, that deviation exceeding ± 1 dB become



Figure 3.18. Modeled auditory deviations of compensated HPTFs from target function for each band of an ERB filter bank for left ears of 25 subjects and six different inversion approaches. Mean deviations of single subjects (grey). Mean deviation across all subjects (red).

audible. However, high Q notches are believed to be perceptively less relevant than remaining high Q peaks. Clear differences between unconstrained and minimum phase compensation, that are probably induced by the filter length, can be found below 200 Hz. As mentioned in the previous section, the unconstrained phase approach is imprecise is this range. In conjunction with generic and individual compensation, this leads to audible boosts of up to 5 dB. However regarding the non-individual inversion, the unconstrained phase approach yields better results than the minimum phase, which can be explained comparing the HPTFs of FABIAN to those of the 25 subjects (see Fig. 3.13). A 5 dB peak between 40 Hz and 200 Hz can be found in the FABIAN HPTFs, which is most likely caused by leakage. This peak is nearly perfectly compensated by the non-individual minimum phase, but not by the unconstrained phase compensation filters. Thus, damping is caused, if the minimum phase filter is applied to the individual HPTFs that show no, or a less pronounced peak within this range. The following discussion of the effect of non-individual, generic and individual headphone compensation is thus restricted to the minimum phase approach.
Non-individual compensation exhibits some leakage-caused damping below 200 Hz, considerable boosts and notches occur between 1 kHz and 5 kHz, whereas increasingly chaotic deviations can be found above 5 kHz. The generic compensation performs better; deviations are symmetrical around the target function up to 5 kHz causing less absolute error. Variance among the subjects remains unaltered. Besides, in comparison to the non-individual approach, the generic compensation reduces the maximum possible error: If HPTF magnitudes across subjects are – at each frequency bin – assumed to be symmetrically distributed, a generic filter based on the average HPTF will always halve the compensation error that would occur between a worst-case pair of individuals when using the non-individual approach. However, high frequency boosting emerges above 5 kHz and ringing artifacts might still be audible. High frequency boosting vanishes almost completely, if individual compensation is applied. Potentially audible deviations below 200 Hz are most likely due to the limited frequency resolution of the FIR-filter. Between 200 Hz and 5 kHz, deviations stay within ± 1 dB. For all methods, compensation results exhibit negative deviations above 6 kHz, due to the high-pass regularization used in the LMS inversion, where the compensation of narrow notches, which are assumed to be barely audible, is avoided. Auditory deviations from the target function are summarized in Tab. 3.6 (Brinkmann and Lindau, 2010).

	50 Hz – 200 Hz	200 Hz – 2 kHz	2 kHz – 5 kHz	5 kHz – 21 kHz
non-individual	+2/-1.7	+1.8 / -1.6	+2.1 / -6.3	+9.1 / -11
	-0.1 / -4.5	+1.8 / -1.8	+2.1 / -6.3	+9.5 / -11.1
generic	+5.1 / -0.3	+1/-1	+3.5 / -3.7	+7.1 / -12.9
	+1.9 / -1.8	+0.9 / -1.1	+3.6 / -3.6	+7/-13.7
individual	+3.8 / +0.2	+0.4 / -0.2	+0/-0.7	+0.7 / -9
	+0.3 / -2	+0 / -1	+0/-0.6	+0.6 / -9.6

 Table 3.6. Modeled auditory deviations of compensated HPTFs from target functions (max/min).

 Differences referring to unconstrained phase compensation are colored.

3.3 Chapter summary

Based on the anatomy of the human ear canal as well as mechanical and electro-acoustical demands, the PRECISE measuring system for recording individual binaural transfer functions was developed. It consists of three differently sized pairs of silicon ear moulds with flush-cast miniature microphones and a *Matlab*[©] based software framework that allows fast HPTF measurement and compensation. The PRECISE ear moulds have shown to fit a large variety

of ear canals and turned out to be a highly reliable measurement instrument.

Further, the effect of six different approaches towards headphone compensation in binaural synthesis was examined. Looking at the linearization of the binaural transmission chain depicted in Fig. 2.3, generic and individual compensation promise noticeable improvements regarding spectral coloration as well as group delay and phase distortion, when beeing compared to the non-individual approach. However, only an individual compensation completely eliminates possibly audible high frequency ringing, whereas perceptively less relevant notches remain uncorrected (Brinkmann and Lindau, 2010).

Chapter 4

Perceptual Evaluation

As part of the present study, two listening tests have been conducted. The first was designed to be a follow up study to Schärer (2008) \Rightarrow 4.1 Listening Test I, and the second emerged from surprising results found in the first test \Rightarrow 4.2 Listening Test II. It was carried out to complete the investigation of the initial research question of the perceptual suitability of different compensation filters. In both tests, the similarity of the binaural simulation compared to a real sound field radiated by a loudspeaker was assessed. The methods and results are discussed in the subsequent sections; a short description is given by Lindau and Brinkmann (2010). All documents that are needed for a reconstruction of the listening tests, including source code of the graphical user interface, audio content, and a documentation of the setup of the convolution engine can be found in the *Listening test I / II* sections of Appendix D.

4.1 Listening Test I

The first listening test that has been conducted, was designed as a follow up study to Schärer (2008). It aimed at assessing the perceptual quality of a binaurally simulated sound field compared to the corresponding sound field radiated by a real loudspeaker. Beside the main aspect, the examination of the headphone compensation filter (factor *filter (3)*), the temporal behavior of the unconstrained and minimum phase inversion (factor *phase (2)*), the integration of a subwoofer (factor *reproduction mode (2)*) and the influence of two stimuli – pink noise and a guitar sample (factor *content (2)*) – were also explored (see Tab. 4.1). This lead to $3 \cdot 2 \cdot 2 \cdot 2 = 24$ conditions that were tested using a fully repeated measures test design.

	Inversion Method	Filter	Phase	Reproduction Mode	Content
Schärer (2008)	high pass reg. LMS Inversion	non-individual	unconstrained	hp only	pink noise, guitar
Present study	high pass reg. LMS Inversion	non-individual, generic, individual	unconstrained, minimum	hp only, hp & subwoofer	pink noise, guitar

Table 4.1. Conditions of Schärer (2008) and present study – listening test I.

Before the listening test was conducted, the following a priori hypotheses were formulated based on the discussions carried out in Chap. 2, 3.2.3 and 3.2.4:

(a) Filter: individual > generic > non-individual

Individual headphone compensation causes the least distortion in the binaural transmission chain and was thus assumed to be perceptively best suited.

(b) **Phase: minimum > unconstrained**

The minimum phase compensation was believed to be better suited, as it reduces the overall system latency and because of results from informal listening test conducted by Norcross et al. (2006).

(c) Reproduction: hp & subwoofer > hp only

Because subjects from Schärer (2008) reported the binaural simulation to suffer from "poor base", the reproduction with an integrated subwoofer was supposed to be perceptively superior.

(d) **Content: guitar > pink noise**

Spectral coloration was reported as the main criterion for the distinction between a real and simulated sound field, when non-individual recordings are used for auralization (Lindau et al., 2007; Schärer, 2008). Since the noise content should reveal the coloration better than the guitar content, the latter was assumed to be rated better.

4.1.1 Design, sample and measure

The listening test was conducted using the *double-blind triple-stimulus with hidden reference* (ABC/HR) test paradigm according to Rec. ITU-R BS.1116-1 (1997). A *Matlab*© based graphical user interface (GUI), developed from code provided by Schärer, was used to acquire the subjects ratings (see fig. 4.1). Button *A* (labeled *Ref*) always triggered the reference. Buttons *B* and *C* (both labeled *Play*) were randomly assigned to the reference and simulation, respectively. Using the three buttons, the subject had to identify the simulation and rate its similarity compared to the reference using the corresponding continuous slider. Only one slider could be moved from its starting position. This way, the slider that had been moved, marked the stimulus that the subject believed to be the simulation. The ratings were then coded by means of difference grades. If the simulation was discriminated (correct answer), negative difference grades ranging from 0 to -4 represented the perceived difference, whereas positive grades in the same range indicated an incorrect answer. In both cases 0 referred to the starting position of the slider and ± 4 to its lower end point.



Figure 4.1. ABC/HR GUI.

Contrary to the ITU recommendation, but following the argumentation of Schärer (2008, pp. 91), the sliders were labeled *identical* and *very different* (german: *identisch* and *sehr unterschiedlich*, respectively). The recommended lables, *imperceptible*; *perceptible*, *but not annoying*; *slightly annoying*; *annoying* and *very annoying*, were discarded because they (a) origin from different semantic categories, (b) have not been shown to be equidistant and (c) do not refer to similarity, which was to be addressed in the current study. No intermediate labels were assigned to ensure equidistance; instead, end and mid-scale markers were provided for better orientation.

Although recommended by the ITU, it was believed that uncompensated control stimuli were not needed, because binaural simulations were clearly detectible in pre-tests. Furthermore, the uncompensated stimuli were by far rated the worst by subjects from Schärer (2008), and it could be argued that this may have masked smaller effects of the remaining conditions. However, the absence of intermediate and extrem control stimuli, as well as the lower end of the scale, which was not anchored, may have lead to different subjective concepts evoked by the label *very different*. This made it impossible to compare the raw difference grades across subjects, and thus, the data was subjected to the z-transform, prior to statistical analysis (see Eq. 3.1).

Six conditions were presented to the subject at a time, meaning that the 24 conditions

were split into four rating sessions. The conditions were randomly assigned, while the content was held constant for two successive sessions to allow better comparability between the conditions. To assure that the obtained ratings are comparable across conditions, the subjects could repeatedly listen to each condition as well as the reference within a session. For comparison with Schärer, the same content was chosen: pink noise to reveal spectral differences between simulation and real sound field and an anechoic classical guitar excerpt to reveal temporal differences. The guitar excerpt has shown to be well suited for revealing the potential flaws of the simulation earlier studies (Lindau et al., 2007; Schärer, 2008). Both had a length of 5 seconds. A drumming excerpt for additional examination of the subwoofer integration was discarded from the test design, because it would have doubled the number of conditions to 48, requiring a larger sample and longer rating times. In any case, it could be argued that the noise content was sufficient for testing the subwoofer integration, as it exhibited enough energy at all frequencies of interest (see Fig. 4.2).



Figure 4.2. Audio content – Magnitude spectrum: pink noise (black), guitar (red) and drumming excerpt (green), 12th octave smoothed.

The desired sample size was derived a priori according to the concept of the optimum sample size, as described by Bortz and Döring (2006, Chap. 9.2.2), using *G*Power 3* for the exact calculation (Faul et al., 2007). The optimum sample size can be calculated for given type-1 error and test power and based on assumptions for effect size and mean inter-subject correlation, i.e. the mean correlation between the ratings across subjects. The effect size specifies the expected differences of mean values between the test conditions. Small effect sizes (E=0.1¹) were assumed and corrected for the factor $1/\sqrt{1-\bar{r}}$, where \bar{r} is the mean inter-subject correlation (Bortz and Döring, 2006, pp. 618). In earlier listening tests, values of $\bar{r} = 0.4$ were observed, which lead to the corrected effect size $E_{corr}=0.1291$. With a desired

¹Effect size index for small effects according to Bortz and Döring (2006, Tab.9.1). This is called f by Faul et al. (2007, Tab. 3)

type-1 error level of α =0.05 and a test power of 1- β =0.8, an optimum sample size of 20 subjects was calculated for a fully repeated measures test design, assuming no interactions¹. It has to be stressed, however, that the calculated sample size is only correct for the assumptions made for inter subject correlation \bar{r} and effect size E. The latter was only estimated though it has a strong influence on the result – for example an effect size of E=0.8 (E_{corr}=0.1033) leads to a sample size of 31.

In total, 27 subjects participated (24 male, 3 female, avg. age 31.7 years), mostly Audio Communication or Sound Engineering students and employes of which 21 had general experiences with listening tests. 24 had a musical background and played an instrument and/or were familiar with music recording and production. Only one subject was neither familiar to listening tests, nor had a musical background. In all, the subjects can be classified as expert listeners as specified in Rec. ITU-R BS.1116-1 (1997). Consequently, the results can not be transferred to a general population, but only hold true for the group of expert listeners that participated in the test.

4.1.2 Setup and validation

The listening test was conducted in a dry studio environment (V = 220 m³, $T_{30 @ 1 \text{ kHz}} = 0.4 \text{ s}$) shown in Fig. 4.3. The subjects were seated on a chair 2 m in front of a Genelec 1031 A loudspeaker, that radiated the real/reference sound field, and used a laptop running the ABC/HR GUI for rating. The simulation was played through electro-static STAX SRS 2050 II headphones for hp only, and through the headphones and an ADAM SUB8 for the hp & subwoofer reproduction mode. The subwoofer was placed underneath the chair and decoupled from the floor with foam underlays to prevent structure-borne sound radiation and vibration. Since the headphones were approximately transparent to exterior sound fields, the simulated and real sound fields could be presented without the need to remove the headphone. Consequently, they remained on each subject's head during the entire listening test, allowing for an instantaneous switching between real sound field and simulation. Before the listening test was conducted, a horizontal BRIR data set (\pm -80° $\leq \phi \leq \pm$ 80°) was measured using the FABIAN HATS, which was placed on the chair in front of the loudspeaker. A spatial resolution of 1° was chosen that had been shown to be inaudible in a previous study (Lindau, 2009a). Elevated BRIR data was not collected, because subjects barely used up/down head movements in earlier studies conducted by the Audio Communication Group. To account for shadowing, diffraction, reflection and absorption effects caused by the headphones, they were placed on FABIAN while the BRIR data set was measured. Afterwards, the ITDs inherent

¹See Appendix B. The number of subjects can be obtained by dividing the total sample size by the number of conditions, in this case $3 \cdot 2 \cdot 2 \cdot 2 = 24$.

to the BRIRs were removed in post-processing and reinserted in real time during the listening test, as described in Lindau et al. (2010a).



Figure 4.3. Setup – listening test I.

Hiekkanen et al. (2009) examined the audibility of subject displacement and horizontal head movements relative to a given stereo loudspeaker setup. For pink noise and binaural recordings made in a standard listening room, displacements of ± 1 cm to the side, 10 cm to the front and a horizontal head movement of $\pm 2.5^{\circ}$, were given as thresholds for audible movements. Regarding the displacement to the side and the head movements, the results equal the spatial resolution used by the experimenter, and thus the actual threshold could well be even smaller. Though the experiment was conducted with a stereo loudspeaker setup, it seems reasonable to expect at least similar thresholds when using a single sound source. Therefore the measurement position of FABIAN, as well as the positions of the subjects were carefully controlled, using the following procedure: The positions of the chair FABIAN and the subjects were seated on, as well as that of the table for the laptop were marked on the floor with adhesive tape. After FABIAN was seated, the height of the loudspeaker was adjusted until the tweeter had the same height has FABIANs interaural axis, and FABIAN was then turned around its vertical axis to face the tweeter. The distance from FABIAN to the tweeter was measured with a laser measurement device. Finally, the position of the interaural axis was determined with two perpendiculars fallen from the studio ceiling. The position of all acoustically effective obstacles, such as tables, effect racks and mixing boards was held constant for the duration of the experiment.

For auralization, the framework described in Chap. 2.2.3 was used. A complete signal flow chart of software and hardware included in the listening test is shown in Fig. 4.4. Two computers were used to carry out the test: a laptop running the ABC/HR GUI randomly controlled the sequence the conditions are presented in, and a second computer managed audio playback and processing. The GUI sent two open sound control (OSC) messages to the rendering computer: one for selecting the desired headphone compensation filter and



Figure 4.4. Signal flow chart – listening test I.

the other for selecting and playing back the audio content. The latter was done using *Pure* $Data^1$, which routed the audio to different outputs, depending on the reproduction mode. For presenting the reference sound field, the audio was routed to output (c). Before it was played back by the loudspeaker, it was filtered with the minimum phase or unconstrained phase target bandpass and with a second bandpass that was applied during the BRIR measurements. If the *hp* only reproduction was chosen, the audio was routed to output (b), where it passed through three processing blocks before being played back through headphones. First, it was convolved with the BRIR corresponding the subjects head position, which was tracked during the entire listening test. The mixing time was set to 140 ms, according to Lindau et al. (2007). Second, the headphone compensation was applied and third, the ITDs were reinserted by means of ITD stretching as described in Lindau et al. (2010a). In the case of *hp* & subwoofer reproduction, the audio stream was routed to outputs (b) and (c). Three additional processing steps were applied, before the audio was played back by the subwoofer. It was convolved with the same BRIR as in output (b) and with the target bandpass. Finally, it was filtered through four parametric equalizers that realized a subwoofer and room compensation, which

¹See http://puredata.info/ (Last checked: July 2011).

was realized with the *Behringer DCX2494* Loudspeaker Controller. The subwoofer integration is described in the next paragraph.

First, the frequency response of the subwoofer was measured in order to determine a cross-over frequency between subwoofer and headphone. Since the anechoic chamber of the Institute for Technical Acoustics and Fluid Mechanics has a lower cut-off frequency of 63 Hz, the frequency response was measured in the near field of the subwoofer, were the influence of room modes is negligible. It has been shown, that the near field of a loudspeaker is proportional to its far field within a tolerance of 1 dB for ka<1 (k: wave number, a: membrane radius) and a distance to the membrane, smaller than 0.11a (Keele Jr., 1974, Eq. 5, 7). Considering the 8" unit of the *SUB8* with a radius of r=10.16 cm, near field measurements should be valid up to about

$$ka < 1 \quad \Leftrightarrow \quad f < \frac{c}{2 \pi a}$$

 $f < \frac{340 \frac{m}{s}}{2 \pi 0.083 m} = 652 \text{ Hz}.$

With the internal low-pass of the *SUB8* set to 150 Hz, the frequency responses of the 8" unit and the bass-reflex port were measured separately and then summed according to Keele Jr. (1974, Eq. 12):

$$L = L_{membrane} + L_{port} + 20 \log_{10} \left(\frac{a_{port}}{a_{membrane}} \right)$$
$$L = L_{membrane} + L_{port} + 20 \log_{10} \left(\frac{3.4 \ cm}{8.3 \ cm} \right)$$
$$L = L_{membrane} + L_{port} - 7.8 \ dB,$$

where L is the level in dB. The resulting magnitude response is depicted in Fig. 4.5(a). Two parametric equalizers have been deployed to achieve -6 dB cut-off frequencies of 24 Hz and 166 Hz, respectively (see Tab. 4.2). Consequently, the lower cut-off frequency of the headphone compensation filter was also set to 166 Hz for the *hp* & *subwoofer* reproduction mode.

In a second step, a room equalization has been applied by means of two parametric equalizers listed in Tab. 4.2, to account for room modes. Lastly, the level and phase of the subwoofer was adjusted to the headphones. The level had been adjusted based on the 1/3rd octave smoothed magnitude response of the subwoofer, the phase could be adjusted by inversion and varying the position of the subwoofer. For verification, a single HPTF has been measured and compensation filters were calculated. Then, the transfer function of the whole electro-acoustical setup, including the signal processing applied by the fWonder



Figure 4.5. Physical validation of test setup. (a) Individually compensated HPTFs measured on experimenters right ear, from top to bottom: (black) hp only reproduction; (red) hp & subwoofer reproduction single channels. (black) near field response of SUB8, equalized (solid) and unequalized (dashed). (Curves offset for clarity, grey lines 1/3rd, colored 1/12th octave smoothed). (b) Corresponding impulse responses (not smoothed).

convolution engines, was measured. If everything had been set up correctly, this would have resulted in a nearly flat frequency response. Results for both, *hp only* and *hp & subwoofer* reproduction mode, are depicted in Fig. 4.5. The magnitude responses show a slight rolloff at high frequencies that vanished when the validation measurement was carried out on the FABIAN HATS, and it was thus assumed that this originateed from small head movements during the measurement process. At frequencies below 166 Hz, a 5 dB ripple, most probably caused by room modes, can be seen for the *hp & subwoofer* reproduction mode (green and red lines, 1/12th octave smoothed). This ripple almost completely disappears if 1/3rd octave smoothing is applied as a rough auditory modeling (grey lines). The impulse responses are shown in Fig. 4.5(b), and a longer decay can be observed for the *hp & subwoofer* reproduction

	Subwoofer equalization	Room equalization		
$\begin{array}{c} \mathbf{E}\mathbf{Q}_1\\ \mathbf{E}\mathbf{Q}_2 \end{array}$	f=40 Hz, G=-6.3 dB, Q=1.4 f=151 Hz, G=6 dB, Q=2	$\begin{array}{c} \mathbf{E}\mathbf{Q}_3\\ \mathbf{E}\mathbf{Q}_4 \end{array}$	f=65 Hz, G=2.7, Q=3.2 f=109 Hz, G=-0.6, Q=1	

mode. Whether or not this is audible had to be examined in the listening test. However, from a physical point of view, the subwoofer integration could be regarded successful.

Table 4.2. Parametric equalizers used for subwoofer and room equalization (Bandwidth definition:
symmetric).

After physical validation, the loudness between simulation and real sound field had to be adjusted. This was done by the experimenter and an expert listener from the Audio Communication Group, based on repeated listening at moderate levels. Subsequently, the volume controls of loudspeaker, subwoofer and headphones were carefully documented. The overall level of both, simulation and real sound field could still be changed in *Pure Data*, if desired. The content was perceptually adjusted as well. However, a perceptual adjustment of the different inversion approaches could not be done, as the individual filters differed across subjects. It was assumed that the normalization at 300 Hz, which was applied to each compensation filter, was sufficient (see Chap. 3.2.2).

4.1.3 Procedure

Before the test started, ten HPTFs and the inter tragus distance, which was needed for ITD stretching, were measured for each subject. While the compensation filters were calculated and the test environment was set up by the experimenter, the subject took his or her time to go through the written instructions (see Appendix B). To familiarize the subject with the rating process, the test started with a training phase, including six preselected exemplary conditions. If no questions arose, the position of the subject was controlled using two perpendiculars as described in Chap. 4.1.2. Before the rating started, the subjects were orally instructed to remain in the established position, encouraged to move their head about the vertical axis and take their time at will for the rating process. Afterwards, a questionnaire was given to the subjects to asses their expertise as well as attributes by which the simulation could be distinguished from the real sound field (see Appendix B). Including inter tragus distance and HPTF measurements, filter calculation and training, the test took 45-60 minutes per subject. On average, 20 minutes were needed for the rating process, which is in accordance to the maximum rating times given in Rec. ITU-R BS.1116-1 (1997).

4.1.4 Analysis and Results

Before subjected to further statistical analysis, a post-screening of the data was carried out to evaluate the expertise of the subjects. Therefore, the ratings were transformed into difference grades, by subtracting the ratings of the simulation from those of the real sound field. This way, negative ratings (0 = identical, -4 = very different) indicate that the simulation has been detected and positive, that the real sound field was believed to be the simulation. Accordingly, only negative difference grades should occur, if the subjects were able to detect the simulation¹. A single sided t-test, carried out separately for each subject, was used to verify this assumption (Rec. ITU-R BS.1116-1, 1997, p. 20). All t-tests showed significant deviations from zero, however, two subjects were excluded from further considerations: One subject encountered technical problems, and the other rated all simulations *very different*. It was thus assumed that the instructions, which explicitly demanded rating the simulations with respect to each other, were misunderstood. In the following, the results based on the difference grades and qualitative judgements of the remaining 25 subjects are described in detail. A post hoc estimation carried out with *G*Power*, showed a power of 0.88 for this sample size (see Appendix B).

Qualitative

The questionnaire was designed to asses the attributes enabling the subjects to distinguish the simulation from the real sound field. The answers were analyzed, assigned to categories and sorted by frequency of occurrence (see Fig. 4.6). Complete answers are given in Appendix B.

In general, the results were well comparable to earlier studies discussed in Chap. 2.2.4. Coloration still was the primary cue and was named to appear especially at high and low frequencies. Both could be caused by the inter-subject variance of binaural signals, the latter in conjunction with leakage (see Chap. 2.5). Only one subject reported the simulation to have "poor bass", on the other hand, two said it had too much bass. This suggests, that results from Schärer (2008) referring this issue could most probably be traced back to a mistake in the test design, where the target bandpass was applied twice to the headphone, but only once to the loudspeaker output. Furthermore, three subjects said, that they made use of loudness cues, which also could have been evoked by strong coloration.

Answers referring to spatiality and source width were provided the second most frequently. Some subjects reported differences in the source width. Other attributes that were given, like "differing reverberation" or "airy simulation" (German: *luftige Simulation*) are hard

¹Only one positive difference grade was found (*content* guitar; *filter* non-individual; *phase* unconstrained; *reproduction mode* hp & subwoofer).



Figure 4.6. Attributes allowing the distinction between real and simulated soundfield – listening test I. Numbers denote relative and (absolute frequencies)

to translate to technical parameters. Moreover, it is not clear, whether or not the subjects differentiated between spatiality and localization, which was also named frequently. Besides a general reference to localization, it was reported that the perceived distance to the sound source differed between simulation and reality, and as discussed in Chap. 2.1, this could be caused by coloration. An unstable or changing localization was also observed by some subjects, which could be explained by mismatching ITDs (Lindau et al., 2010a). Sometimes, a blurry or indefinite localization has been reported as well, which also could refer to the perceived source width and support the assumption, that some subjects did not distinguish between localization and spatiality.

Furthermore, two subjects reported differences in the temporal behavior that might indicate an audible difference between minimum phase and unconstrained phase inversion, two reported that they were able to detect the subwoofer, and one subject experienced missing externalization at low frequencies. However, how much the qualitative judgements are reflected by the ratings has to be revealed by further statistical analysis.

Inferential

Before being analyzed by means of a four-factorial analysis of variance (ANOVA), some descriptive values have been considered. For estimating the reliability of the results, Cronbach's alpha parameter was calculated. It is a measure for the internal consistency of the data and increases with growing inter-item correlation (Bortz and Döring, 2006, pp. 198). A relatively high value of Cronbachs's alpha = 0.944 was found, indicating that the ratings across subjects are in good accordance to each other. Second, the distribution of the difference grades for each conditions was tested for normality (Lilliefors, 1967). Eight out of the 24 conditions were clearly normaly distributed, of whom six distributions corresponding to the guitar content all exhibited an inclination to the right, showing that the majority of subjects rated them to be rather similar to the reference (see Appendix B).

The mean difference grades and 95% confidence intervals for all 24 conditions are depicted in fig. 4.9. The negative ratings show, that the simulation was clearly detectable. Moreover,



Figure 4.7. Results of listening test I: Mean differences grades and 95% confidence intervals across conditions, averaged over all subjects. Factor filter is indicated by colored areas, factors phase and reproduction mode by arrows.

an obvious difference in the ratings of noise and guitar content, and a trend for non-individual filters to be rated better than generic and individual filters can be observed. Differences in phase and reproduction mode are less obvious. As mentioned before, z-scores of the rating were calculated before ANOVA, as no anchor stimuli were integrated in the test design. The inference statistical analysis was carried out in *SPSS*, the post-screening including the z-score calculation has been done in *Matlab*©. The only violation of preliminary ANOVA conditions was the missing sphericity for the filter*content interaction (Mauchly's W = 0.748, p = 0.035; see Mauchly, 1940). The main effects and significant interactions are listed in Tab. 4.3, a short summary of the statistical output is given in Appendix B, and *SPSS* files containing the

Factor	df	F	р	Part. Eta squared	Observed power
Content	1	45.055	0**	0.652	1
Filter	2	19.848	0**	0.453	1
Phase	1	2.146	0.156	0.082	0.29
Reproduction	1	2.008	0.169	0.077	0.275
Phase*reproduction	1	9.960	0.004**	0.293	0.857
Filter*phase*reproduction	2	4.536	0.016**	0.159	0.746

ratings, as well as syntax and output data can be found in Appendix D.

Table 4.3. Listening test I: Main effects and significant interactions.(Asterisks mark significance, sphericity assumed)

Two highly significant main effects can be observed: the noise was rated worse than the guitar content and second, non-individual was rated better than both, generic and individual headphone compensation, as was revealed by pairwise Bonferroni post hoc tests. The test for differences between generic and individual compensation showed no significant result, but a trend for the latter being rated worse could be seen from the mean values. Further, no significant results were found for the remaining factors *reproduction mode* and *phase*, however, a trend for the unconstrained being rated slightly worse than the minimum phase inversion was found.

Besides this, two highly significant interactions were found: ratings across *reproduction modes* are similar for minimum, but differ for unconstrained phase inversion, where the hp only reproduction is rated worse than the hp & sub (see Fig. 4.8(a)). This interaction seemed to influence the factor *filter* as well (see Fig. 4.8(b) and 4.8(c)). While ratings for non-individual compensations are nearly identical, generic and individual filters are rated worse for hp only reproduction in conjunction with unconstrained phase.

4.1.5 Discussion

The effect of *content* is according to the a priori hypothesis and rather obvious. The coloration of the simulation, can be detected better with a noise stimulus, because it (a) is a stationary signal and (b) has more energy in at the affected low and high frequency ranges than the guitar stimulus (see Fig. 4.2). The effect of the compensation filter is surprising and contrary to the a priori hypotheses. Because the underlying mechanism is complex, it will be discussed in a separate section. The hypotheses regarding *phase* and *reproduction mode* could not be verified and consequently the subwoofer integration can be regarded as successful. As indicated by the qualitative analysis, this however does not mean, that the hp & subwoofer



Figure 4.8. Significant interactions – Listening test I. (a) phase*reproduction mode, (b, c) filter*phase*reproduction mode

reproduction could not be distinguished from the hp only reproduction, but only that not even small effects occurred. In general, the same holds true for the phase behavior, but the interaction between phase and reproduction mode has to be considered as well. The low ratings for the hp only condition in conjunction with the unconstrained phase response could possibly be explained by its insufficient low frequency compensation. As shown in the left column of Fig. 3.18¹, unconstrained phase inversion leads to a bass emphasis and some subjects reported either a rumbling bass, or perceived the simulation to be dull compared to the real sound field (German: *grummeln im Bass* and *dumpf*, respectively). As the subwoofer is not affected by this, it possibly explains the better ratings for hp & subwoofer reproduction, which are comparable to the minimum phase inversion.

Furthermore, this consideration already gives a trace for a possible explanation to the second order interaction between content, phase and reproduction mode. As depicted in Fig. 3.18, no bass emphasis can be seen for the unconstrained phase non-individual compensation. As discussed in Chap. 3.2.4, this is caused by an interaction between FABIAN HPTFs and and unconstrained phase filters. Combining these aspects, the second order interaction could be explained, suggesting that in this case, a slight bass emphasis was perceptually more salient than a loss.

Effect of compensation filters

To assess the effect of the filter type (non-individual, generic, individual), the corresponding signal transmission chains have to be analyzed. Regarding the real sound field, the complete transmission is described by the BRIRs of the listener. When individual headphone compen-

¹Or Fig. 3.14 – 3.16 for a more detailed presentation, that is restricted to a single subject.

sation is applied, the transmission is described by FABIAN's BRIRs, since the HPTF of the listener is completely compensated¹. Consequently, a comparison between real sound field and simulation is equivalent to a comparison of the listeners BRIRs to foreign BRIRs, in this case belonging to FABIAN. If the headphones are compensated with non-individual filters, the foreign BRIRs are "colored" by the filters, and these colored BRIRs are then compared to those of the listener.



Figure 4.9. Effect of compensation filter. (a) BRIR of FABIAN (black) and mean BRIR of 5 subjects(red). (b) Differences from binaural simulation to individual BRIR. Non-individual BRIR and headphone compensation (black), non-individual BRIR and individual headphone compensation (red). Mean curves of 5 subjects (All curves averaged over both ears and 3rd octave smoothed).

For a physical description of these transfer functions, HPTFs and BRIRs for frontal head orientation ($\phi = \theta = 0^{\circ}$) of FABIAN and five subjects have been measured in an office room. In a next step, the FABIAN BRIRs have been filtered with individual and non-individual headphone filters and the difference to the subjects BRIRs was calculated. The results, averaged over both ears and 5 subjects, are given in Fig. 4.9(b). Coloration is mainly introduced in the region above 5 kHz, and indeed, less damping is found for the nonindividual headphone compensation. In both cases, individual and non-individual, the

¹Disregarding measurement errors and regularization for simplicity.

damping originates from the differences in the BRIRs that is depicted in Fig. 4.9(a), but somehow, the non-individual headphone compensation seems to reduce the influence of the foreign BRIR. Similar results can be observed for the generic compensation, but are not depicted due to simplicity.

A possible explanation for this might be, that prominent pinna features of FABIAN's near field HRTFs are maintained in the HPTF, especially at high frequencies and when measuring with circumaural headphones (Møller et al., 1995c, p. 314, Kulkarni and Colburn, 2000, Lindau and Brinkmann, 2010). If these pinna features could as well be found in FABIAN's BRIRs, which is likely as they can be understood as a superposition of HRTFs, the non-individual compensation could have lead to some kind of deindividualization of FABIAN's BRIRs. This could explain the results observed in the listening test, however, the considerations are based on only five subjects.

An informal listening test was conducted to support this rather theoretical approach. The differences depicted in fig. 4.9(b) were auralized through headphones by means of filtered pink noise and compared to the unfiltered equivalent. Results from this test confirmed the findings, however, further investigations could be carried out, examining whether or not this effect can be observed for different source locations, head orientations and non-individual BRIRs not taken from the FABIAN HATS, but arbitrary human subjects.

4.2 Listening Test II

As results from the first listening test regarding the effect of *filter* were unexpected, a second listening test was conducted. This time, individual headphone compensation was compared to true non-individual compensation, based on HPTFs from a randomly chosen human subject (factor *filter (2)*). Further, the influence of the phase response was tested again, because an interaction between *phase* and *reproduction mode* occurred in the first listening test (factor *phase (2)*). Besides pink noise, a drumming excerpt was chosen, as its prominent transient components should reveal any insufficient temporal behavior (factor *content (2)*). In addition, three different regularization functions were tested since the high-pass regularization – or high shelve to be more precisely – causes a damping in the compensated HPTFs (factor *regularization (3)*). The newly introduced regularization function are described in Chap. 4.2.1. In total, this lead to $3 \cdot 2 \cdot 2 \cdot 2 = 24$ conditions, that were tested using a fully repeated measures test design (see Tab. 4.4).

Inversion Method	Regularization	Filter	Phase	Content
LMS Inversion	high-pass, inv. HPTF, PEQ	true non-individual, individual	unconstrained, minimum	pink noise drum

Table 4.4. Conditions of listening test II.

Again, hypothesis were made regarding the perceptual suitability of the factor levels:

(a) **Regularization: PEQ > inverse HPTF > high-pass**

Physical evaluation showed best results for PEQ regularization regarding damping at high frequencies (see Chap. 4.2.1).

(b) Filter: individual > true non-individual

The true non-individual headphone filter introduces unwanted coloration and was therefore believed to be perceptively inferior. The amount of coloration should be comparable to that of non-individual compensation, however, it should not be related to the binaural recordings used for auralization.

(c) Phase: minimum > unconstrained

Although no effect could be observed in the first listening test, minimum phase compensation was still believed to be better suited, as it reduces the overall system latency, and because of results from informal listening test conducted by Norcross et al. (2006).

(d) Content: drum > pink noise

Spectral coloration was reported as the main criterion for the distinction between a real and simulated sound field, if non-individual recordings are used for auralization (see Chap. 4.1.4). Since the noise content should reveal the coloration better than the drumming excerpt, the latter was assumed to be perceptually better suited.

The second listening test was as well conducted in a dry studio environment (V = 145 m³, $T_{30 \ @\ 1 \ kHz}$ = 0.47 s). Since the test design, measure, setup, procedure and analysis were nearly identical to listening test I, refer to Chap. 4.1 for their description. The design only differed regarding the new conditions, and the setup was identical besides the subwoofer, which was not used in the second test.

4.2.1 Possible compensation improvements

The second listening test was taken as an opportunity for further refinement of the compensation method, as it was seen, that in average, the high-pass regularization caused damping at high frequencies in the compensated HPTFs, when individual compensation was applied (see Fig. 3.18). Therefore, five new methods were implemented. Since it was assumed that the HPTFs should be perfectly equalized within the whole frequency range, except for the frequencies, where notches occurred, the new methods allowed for a more selectively regularization. All used LMS inversion, but different approaches to regularization. They are described in the following. For clarity, an example is given for each method, showing 10 HPTFs (grey), the compensation filter (black) and the regularization function (red) on the left hand, and the computed compensation result on the right hand side.

(a) High-pass regularization

As used in the first listening test and described in Chap. 2.4.



(b) Parametric equalizer regularization

The regularization function was manually composed from 1 to 3 parametric equalizers, whose center frequencies matched with the notches in the HPTFs. This way, regularization was only applied to notches. If the HPTFs exhibited sharp notches, high Q's and gains were used (30 < Q < 50, G=15 dB), otherwise moderate settings were chosen (1 < Q < 5, G<5 dB). The beta weight was always set to β =1. After the regularization function was composed, a hard limiter could be used in order to obtain plateaus with equal gain and thus equal regularization. In a final step, the regularization function was rescaled to a dynamic range of 20 dB. This way, regularization was very flexible and could be adapted to the HPTFs depending on notches and intra-individual variance.



(c) Inverse and smoothed HPTF regularization

The directly inverted and 1/6th octave smoothed magnitude response of the average HPTF was used for regularization. The HPTF was set to a constant below and above given limits before it was inverted. This way, excessive gains in the regularization function that would have dominated the regualrization, close to 0 Hz and the Nyquist frequency, were avoided. The limits were in almost every case set to 300 Hz and 20 kHz, respectively. Further, the inversion dynamic could be restricted, to avoid excessive gains between the given limits, but this was rarely needed. If inversion dynamic was applied, the inverted regularization function was clipped if it exceeded a threshold that was defined with respect to the minimum value. In most cases, good results were obtained with beta weights of β =0.2. On one hand, this approach worked good with the default parameters in most cases, on the other there are little possibilities for adjustment. A similar approach was used by Schärer (2008), however she used broader smoothing of one octave that flattened the regularization function and thus made it less selective.



(d) Inverse and selectively smoothed HPTF regularization

This approach is similar to (c), but in this case the 1/9th octave smoothed and averaged magnitude response of the HPTF is subtracted from the octave smoothed magnitude response on a dB basis. This replaced the inversion carried out in (c). Although the resulting regularization functions were similar to approach (c), stronger peaks but softer dips were found for this approach, allowing for a more selectively regularization. The HPTF was again set to a constant below and above given limits to avoid excessive gains. Inversion dynamic was not implemented, as it was not needed with this approach, but could easily be added. This approach as well worked good with the default parameters, but leaves little room for adjustment. It is comparable to the *compare and squeeze* method introduced by Müller (1999)



(e) Inversion of smoothed and averaged HPTF, no regularization

Instead of the averaged HPTF in conjunction with a regularization function, the smoothed and averaged HPTF without regularization was subjected to LMS inversion. For this approach, 1/6th octave smoothing was chosen. This represented a good compromise between broader smoothing that would have lead to an imprecise compensation and finer smoothing, where notches were maintained and compensated as well. After the smoothing width was chosen, a fully automated headphone compensation is obtained with this approach without any possibilities for adjustment.



(f) Inversion of smoothed an averaged HPTF with limited inversion dynamic, no regularization

Identical to approach (e), but in addition the inversion dynamic could be limited to avoid excessive gains as in approach (c). This however, was barely needed.



Since not all of the new inversion approaches could be examined in the listening test, an auditory modeling, as described in Chap. 3.2.4 was done and only the two most promising new approaches were selected. The results are shown in Fig. 4.10. Compared to high-pass regularization, less damping at high frequencies is caused in any case. The direct inversion of the smoothed HPTF – approaches (e) and (f) – was discarded, because high deviations from the target function of up to +9 dB were found above 10 kHz, which might have evoked audible ringing artifacts. Results for smoothed HPTF regularization – approaches (c) and (d) – are very similar, however, as bigger deviations are found for the selectively smoothed HPTF approach, it was discarded as well. The remaining compensation methods both were promising: PEQ regularization showed comparably small damping at high frequencies while positive deviations are much smaller than for approaches (e) and (f). They were below +2 dB up to 17 kHz. Using inverse HPTF regularization, nearly perfect compensation could be obtained up to 10 kHz, which was unique to all considered approaches. Above 10 kHz, deviations were below +2 dB.



Figure 4.10. Modeled auditory deviations of compensated HPTFs from target function for each band of an ERB filter bank. Six different inversion approaches, individual headphone compensation. Mean deviations of 25 subjects (grey). Mean deviation across all subjects (red).

4.2.2 Sample

In total, 29 subjects attended (22 male, 7 female, avg. age 27.55 years), again mostly Audio Communication or Sound Engineering students and employes of which 23 had general experiences with listening tests. 23 had a musical background and played an instrument and/or were familiar with music recording and production. Only two subject were neither familiar to listening test, nor had a musical background. In all, the subjects could be classified as expert listeners as demanded in Rec. ITU-R BS.1116-1 (1997). Consequently, the results can not be transferred to a general population, but only hold true for the group of expert listeners that participated in the test.

4.2.3 Results

A single sided t-test, carried out separately for each subject, was used for post-screening (Rec. ITU-R BS.1116-1, 1997, p. 20). All t-tests showed significant deviations from zero, however, six positive difference grades were found in the data, two in conjunction with noise, four with the drumming excerpt. Two subjects were excluded from further considerations because of technical problems. In the following, the results based on the difference grades and qualitative judgements of the remaining 27 subjects are described in detail. A post hoc estimation, corrected for the observed mean inter subject correlation, carried out with G*Power, showed a power of $1 - \beta = 0.93$ for this sample size (see Appendix B).

Qualitative

The attributes that enabled the subjects to distinguish the simulation from the real sound field were analyzed, assigned to categories and sorted by frequency of occurrence (see Fig. 4.11). Complete answers are given in Appendix B.



Figure 4.11. Attributes allowing the distinction between real and simulated soundfield – listening test II. Numbers denote relative and (absolute frequencies)

The results were well comparable to the first listening test, as the order of the categories was identical, apart from reversed appearance of *missing externalization* and *temporal course*¹. The only apparent difference compared to listening test I was found for coloration, which was mentioned by every subject. In addition, ringing artifacts were reported frequently, possibly

 $^{^{1}\}chi^{2}$ tests for significant differences in absolute frequencies between listening test I and II were not calculated.

evoked by true non-individual compensation methods. Further, some subject said that they perceived the timbre of the simulation to be "broken" or "horrible" (German: *kaputt* and *grausam*, respectively) or observed serious phase artifacts. Such strongly negative attributes were not reported in the first listening test. For a more detailed discussion of the remaining attributes, refer to Chap. 4.1.4

Inferential

Again, inter-rater reliability was high (Cronbach's α =0.919) and tests for normality of the ratings showed that normal distribution could not be assumed for nine conditions. Seven of them exhibited an inclination to the right, showing that the majority of subjects rated them to be rather similar to the reference. They all referred to the drumming excerpt (see Appendix B). Mauchly's test for sphericity showed no violations of ANOVA preconditions. The mean difference grades and 95% confidence intervals for all 24 conditions are depicted in Fig. 4.12. The negative ratings indicate that the simulation was clearly detectable. Further,



Figure 4.12. Results of listening test II: Mean differences grades and 95% confidence intervals across conditions, averaged over all subjects. Factor filter is indicated by colored areas, factors phase and regularization by arrows.

a clear difference in the ratings for content, with noise being preceptively more critical, was seen. Moreover, the true non-individual filters were in average rated worse than the

individual filters. Differences between different regularization or phase response approaches were less obvious.

These obvious differences in the ratings were supported by ANOVA, revealing two highly significant main effects for content and filter (see Tab. 4.5). Further, a significant interaction between filter and regularization was observed, where the rating order between regularizations approaches was reversed for individual compared to true non-individual headphone compensation. As depicted in Fig. 4.13, inverse HPTF regularization was rated best for individual, but worst for true non individual compensation. High-pass regularization was rated second best for both cases, and PEQ regularization worst for individual, but best for true non-individual compensation.

Factor	df	F	р	Part. Eta squared	Observed power
Content Filter Phase Regularization	1 1 1 2	55.017 110.765 0.000 0.834	0** 0** 0.983 0.441	0.679 0.810 0.000 0.031	1 1 0.050 0.185
Filter*regularization	2	10.496	0.000**	0.288	0.984





Figure 4.13. Significant interaction – Listening test II: Filter*regularization.

4.2.4 Discussion

The effects found for *content* and *filter* were in accordance with the a priori hypothesis and a possible explanation is rather obvious. Since coloration was the main cue for the detection of the simulation, the noise content was rated worse, as it was more critical to this. The effect

regarding the factor *filter*, can possibly be explained by the coloration that was induced by the true non-individual HPTF compensation. The interaction between *filter* and *regularization* suggests, that inverse HPTF regularization is perceptually best suited, if individual headphone compensation is applied. However, a plausible explanation for the entire interaction can not be given so far. The superiority of the inverse HPTF regularization in conjunction with individual compensation could possibly be explained by the nearly perfect compensation up to 10 kHz (See Fig. 4.10). On the other hand, this might emphasize differences between individual and (true) non-individual HPTFs and induce stronger coloration. Hence, poorer ratings might be obtained for the true non-individual conditions. Further, it could be argued that the PEQ regularization is best for preventing the inversion of notches, since heavy regularization was applied (β =1) and it could be well adapted to the position of the notches. Hence audible ringing artifacts in the region of the first notch (7–8 kHz) are unlikely to appear regardless of the filter that is used. This could possibly explain the good ratings obtained for true non-individual compensation, as ringing artifacts were frequently named to degrade the simulation. However, this does not explain poor ratings for individual headphone compensation. Here, notches that are inherent to the HPTFs are simply maintained by PEQ

regualarization, which should not lead to a degradation of the binaural simulation, with respect to inverse HPTF and high-pass regularization. Double notches, that were mentioned by Norcross et al. (2006), should not have been evoked in conjunction with individual compensation, since the regularization was nearly perfectly adapted to the HPTFs.

4.3 Chapter summary

Two listening test were conducted which showed that the influence of the compensation filter phase response is negligible. Further, it was proven that headphones were well capable of radiating the desired sound field at moderate listening levels and with the lower end of the transmission range restricted to 50 Hz. In addition, a subwoofer can be integrated into the binaural simulation without considerable degradation, which could be compelling if the mentioned restrictions are not wanted. Surprisingly, non-individual headphone compensation turned out to be perceptively best suited, when non individual recordings and HPTFs were measured on the same HATS. It was assumed that an interaction with the non-individual BRIRs caused this outcome. Generic and individual headphone compensation were rated worse than non-individual but by far better than true non-individual compensation, which should be avoided in any case.

Chapter 5

Conclusion

By means of binaural synthesis, virtual acoustic environments can be created by recording and reproducing the signals arriving at a listeners ears. In theory, an auditory scene can be perfectly reproduced if headphones – compensated for the headphone transfer function – are used for the reproduction of the binaural recordings. However, besides possible coloration induced by the headphones, binaural recordings also introduce strong coloration, when they are not made on the listeners ears. As this most often is the case, coloration was reported to be the main cue that enables the listener to distinguish between a real and a simulated acoustical environment. Therefore, the perceptual suitability of different approaches towards headphone compensation was examined, with a focus on non-individual binaural recordings.

5.1 Summary

Since individual HPTFs had to be measured in order to achieve individual headphone compensation, a measuring instrument for recording at the blocked ear canal was developed. Based on measures of the human ear canal, silicon ear moulds with flush-cast miniature electret condenser microphones were crafted in three different sizes. It was shown that the ear moulds fit a large variety of ear canals and that they are a highly reliable tool for fast in situ HPTF measurements. Consequently, the measuring system was named PRECISE (precisely repeatable acquisition of individual headphone transfer functions). To obtain a first impression of the effect of headphone compensation, a physical evaluation was done. Therefore, individual HPTFs were measured and compensated with non-individual, individual and generic compensation filters. The latter were obtained from the average HPTF of 25 subjects. For compensation, the so called LMS inversion with high-pass/high-shelve regularization was used. The least coloration was induced by individual headphone compensation, where only high Q notches remained uncompensated, whereas heavy coloration was induced by both, generic and non-individual compensation.

Since the physical evaluation only comprised the headphone transfer function, but not the binaural room impulse responses, a perceptual evaluation was done as well. Therefore, two listening tests were conducted, where the similarity between a real sound field, radiated by a loudspeaker, and the binaural simulation, based on recordings with the FABIAN HATS, was assessed. Surprisingly, non-individual headphone compensation, based on HPTFs of FABIAN,

turned out to be perceptually best suited. It was hypothesized that this was caused by an interaction between FABIAN's HPTFs and BRIRs: as the HPTFs resemble prominent features of near field HRTFs, which again can be found in BRIRs, some kind of de-individualization was assumed. This hypothesis could be supported by informal measurements and listening tests. Furthermore, no significant difference was found between individual and generic compensation, however, a true non-individual compensation, based on HPTFs from a randomly chosen subject, showed to be perceptually inferior to all, individual, generic and non-individual compensation. Hence, a true non-individual compensation has to be avoided in any case. A non-individual compensation has to be preferred to individual compensation in case of matching BRIRs and HPTFs, at least in conjunction with FABIAN.

Besides the origin of the compensation filter, the perceptual influence of a minimum phase and linear/unconstrained phase inversion was examined. Both phase responses exhibited possible perceptual disadvantages: the minimum phase compensation introduced group delay distortion, the unconstrained phase compensation introduced preringing. However, the phase response showed to have no effect, despite highly critical content – a drumming excerpt – was chosen. Thus, the minimum-phase response should be preferred, as better results regarding the desired magnitude response can be obtained for a given filter order.

In addition, it was shown that a subwoofer can be integrated into the simulation without perceptual degradation. Although headphones were capable of radiating the desired sound field at moderate listening levels and with a restriction to the lower end of the magnitude response, this possibility might be compelling for other applications.

5.2 Perspectives

Regarding the desired authentic binaural reference system, two improvements of the binaural simulation have to be considered in particular: First, the PRECISE ear moulds could be used for recording individual BRIRs and second, the perceptual effect of extraaural headphones developed by Schultz (2011) has to be examined. The extraaural headphones promise less intra-individual variance in HPTFs and better agreement with the FEC criterion. They further don't have to be removed from the subjects head, between HPTF measurement and filter calculation. Consequently, if individual BRIRs were used for auralization, only measuring errors regarding (a) the position of the PRECISE ear moulds at the blocked ear canal and (b) the measurement process of individual BRIRs would remain. Whether these errors lead to a perceptual degradation of the binaural simulation, or an authentic simulation can be obtained will be subjected in future studies. Since the measurement of individual BRIR data sets would be time consuming, algorithms for interpolation could be investigated as well.

Another point would be the examination of the de-individualization effect, that was found in the first listening test. Here it is of interest, whether or not this effect can be observed for different source positions, head orientations and other BRIRs than that of FABIAN. Further the integration of the subwoofer could be improved by compensating for the transfer function from subwoofer to ear drum, if necessary. Moreover, it could be examined whether or not the repositioning of the PRECISE ear moulds leads to audible artifacts. Therefore, a listening test comparable to Paquier and Koehl (2010) could be conducted where multiple binaural recordings would be made for repositioned ear moulds. Lastly, the PRECISE ear moulds could be considered as an actual microphone for artistical use.

Bibliography

- Ahnert W and Tennhardt H P (2008): "Raumakustik". In: S Weinzierl (Ed.) *Handbuch der Audiotechnik*, 1st ed. Berlin Heidelberg: Springer, pp. 181–266. (Cited on p. 5.)
- Alexander M (1968): "Anthropometry of the human ear. (A photogrammetric study of USAF flight personnel)". Tech. rep., Aerospace Medical Research Laboratories, Wright-Patterson Air Force Base, Ohio, USA. (Cited on p. 32.)
- Algazi V R, Avendano C and Thompson D (1999): "Dependence of subject and measurement position in binaural signal acquisition". In: *J. Audio Eng. Soc.* 47(11):937–947. (Cited on p. 16.)
- Algazi V R, Duda R O, Thompson D M and Avendano C (2001): "The CIPIC HRTF database". In: *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*. New Paltz, New York, pp. 99–102. (Cited on pp. 24, 32, 34.)
- Blauert J (1997): *Spatial Hearing. The psychophysics of human sound localization*. Revised ed. Camebridge, Massachusetts: MIT Press. (Cited on pp. 3, 4, 5, 13, 25, 32, 33, 35.)
- Blauert J and Laws P (1978): "Group delay distortions in electroacoustical systems". In: *J. Acoust. Soc. Am.* 63(5):1478–1483. (Cited on pp. 22, 55.)
- Blauert J, Mellert V, Platte H J, Laws P, Hudde H, Scherer P, Poulsen T, Gottlob D and Plenge G (1978): "Wissenschaftliche Grundlagen der kopfbezogenen Stereofonie. Bericht über das Vorkolloquium zur DAGA '78 in Bochum, 13. März 1978". In: *Rundfunktech. Mitteilungen* 22(4):195–218. (Cited on pp. 16, 31.)
- Bortz J (2005): *Statistik für Human- und Sozialwissenschaftler*. 6th ed. Heidelberg, Germany: Springer. (Cited on p. 34.)
- Bortz J and Döring N (2006): Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler. 4th ed. Heidelberg, Germany: Springer. (Cited on pp. 66, 74.)
- Brinkmann F and Lindau A (2010): "On the effect of individual headphone compensation in binaural synthesis". In: *Fortschritte der Akustik: Tagungsband d. 36. DAGA*. Berlin, Germany, pp. 1055–1056. (Cited on pp. 31, 51, 61, 62.)
- Burkhard M and Sachs R (1975): "Anthropometric manikin for acoustic research". In: J. *Acoust. Soc. Am.* 58(1):214–222. (Cited on p. 32.)

- Chan J C K and Geisler C D (1990): "Estimation of eardrum acoustic pressure and of ear canal length from remote points in the ear canal". In: *J. Acoust. Soc. Am.* 87(3):1237–1247. (Cited on pp. 32, 35.)
- Dillon H (1977): "Effect of leakage on the low-frequency calibration of supraaural head-phones". In: *J. Acoust. Soc. Am.* 61(5):1383–1386. (Cited on pp. 25, 51.)
- Eckel G (2001): "Immersive audio-augmented environments: the LISTEN project". In: *Proceedings fifth international conference on information visualisation*, vol. 128. IEEE Comput. Soc., pp. 571–573. (Cited on p. 15.)
- Faul F, Erdfelder E, Lang A G and Buchner A (2007): "G*Power 3: A felxible statistical power analysis program for the social, behavioral, and biomedical sciences". In: *Behavior Research Methods* 39(2):175–191. (Cited on pp. 66, 118, 125.)
- Ford A (1942): "The binaural intensity disparity limen". In: *J. Acoust. Soc. Am.* 13:367–372. (Cited on p. 4.)
- Hammershøi D and Møller H (1991): "Free field sound transmission to the external ear. A model and some measurements". In: *Fortschritte der Akustik DAGA 1991*. pp. 473–476. (Cited on p. 6.)
- Hammershøi D and Møller H (1996): "Sound transmission to and within the human ear canal". In: *J. Acoust. Soc. Am.* 100(1):408–427. (Cited on pp. 6, 14.)
- Hiekkanen T, Mäkivirta A and Karjalainen M (2009): "Virtualized listeneing tests for loud-speakers". In: *J. Audio Eng. Soc.* 57(4):237–251. (Cited on pp. 30, 68.)
- IEC 268-13 (1985): "Sound system equipment, Pt. 13, listening tests and loudspeakers". Tech. rep., International Electrotechnical Commission. Geneva, Switzerland. (Cited on p. 23.)
- Keele Jr D B (1974): "Low-frequency loudspeaker assessment by nearfield sound-pressure measurement". In: *J. Audio Eng. Soc.* 22(3):154–162. (Cited on p. 70.)
- Kirkeby O and Nelson P A (1999): "Digital filter design for inversion problems in sound reproduction". In: *J. Audio Eng. Soc.* 47(7/8):583–595. (Cited on p. 18.)
- Kirkeby O, Nelson P A, Hamada H and Orduna-Bustamante F (1998): "Fast deconvolution of multichannel systems using regularization". In: *IEEE Trans. on speech and audio processing* 6(2):189–194. (Cited on p. 18.)
- Kleiner M, Dalenbäck B I and Svensson P (1993): "Aurlization An overview". In: *J. Audio Eng. Soc.* 41(11):861–875. (Cited on p. 10.)
- Klemm O (1920): "Untersuchung über die Lokalisation von Schallreizen IV: Über den Einfluss des binauralen Zeitunterschieds auf die Lokalisation". In: *Arch. ges. Psychol.* 40:117–145. (Cited on p. 4.)
- Kulkarni A and Colburn H (2000): "Variability in the characterization of the headphone transfer function". In: *J. Acoust. Soc. Am.* 107(2):1071–1074. (Cited on pp. 20, 29, 51, 79.)
- Lentz T, Assenmacher I and Sokoll J (2005): "Performance of spatial audio using dynamic cross-talk cancellation". In: *119th AES Convention, Convention Paper 6541*. New York, USA. (Cited on p. 6.)
- Lilliefors H W (1967): "On the Kologorov-Smirnov test for normality with mean and variance unknown". In: *Journal of the American Statistical Association* 62(318):399–402. (Cited on p. 75.)
- Lindau A (2006): *Ein Instrument zur softwaregestützten Messung binauraler Raumimpulsantworten in mehreren Freiheitsgraden*. Magisterarbeit, Technische Universität, Berlin, Germany. (Cited on pp. iii, 10, 40.)
- Lindau A (2009a): "On the spatial resolution of virtual acoustic environments for head movements on horizontal, vertical and lateral direction". In: *EAA Symposium on Auralization*. Espoo, Finnland. (Cited on p. 67.)
- Lindau A (2009b): "The perception of system latency in dynamic binaural synthesis". In: *Fortschritte der Akustik: Tagungsband d. 35. DAGA*. Rotterdam, Netherland, pp. 1063–1066. (Cited on pp. 11, 13.)
- Lindau A (2010): *Anleitung zum Messroboter FABIAN. (vs. 0.702/02/2010)*. Technische Universität Berlin, Berlin. (Cited on p. 49.)
- Lindau A and Brinkmann F (2010): "Perceptual evaluation of individual headphone compensation in binaural synthesis based on non-individual recordings". In: *3rd ISCA/DEGA Tutorial and Research Workshop on Perceptual Quality of Systems* :137–142. (Cited on pp. 63, 79.)
- Lindau A, Estrella J and Weinzierl S (2010a): "Individualization of dynamic binaural synthesis by real time manipulation of the ITD". In: *128th AES Convention, Convention Paper*. London, UK. (Cited on pp. 11, 27, 68, 69, 74.)

- Lindau A, Hohn T and Weinzierl S (2007): "Binaural resynthesis for comparative studies of acoustical environments". In: *122th AES Convention*. Vienna, Austria. (Cited on pp. 11, 12, 24, 27, 64, 66, 69.)
- Lindau A, Kosanke L and Weinzierl S (2010b): "Perceptual evaluation of physical predictors of the mixing time in binaural room impulse responses". In: *128th AES Convention, Convention Paper*. London, UK. (Cited on p. 11.)
- Lindau A and Weinzierl S (2011): "Assessing the plausibility of virtual acoustic environments". In: *Forum Acusticum 2011*. Aalborg, Denmark. (Cited on p. 1.)
- Lorho G (2009): "Subjective Evaluation of Headphone Target Frequency Responses". In: *126th AES Convention, Convention Paper 7770*. Munich, Germany. (Cited on p. 25.)
- Mauchly J W (1940): "Significance test for sphericity of a normal n-variate distribution". In: *The Annals of Mathematical Statistics* 11(2):204–209. (Cited on p. 75.)
- Menzel D, Wittek G, Theile G and Fastl H (2005): "The binaural sky: A virtual headphone for binaural room synthesis". In: *Tonmeistersymposium*. Hohenkammer, Germany. (Cited on p. 6.)
- Meyer J (2008): "Musikalische Akustik". In: S Weinzierl (Ed.) *Handbuch der Audiotechnik*, 1st ed. Berlin Heidelberg: Springer, pp. 123–180. (Cited on pp. 10, 30.)
- Middlebrooks J C, Makous J C and Green D M (1989): "Directional sensitivity of sound pressure levels in the human ear canal". In: *J. Acoust. Soc. Am.* 86(1):89–108. (Cited on pp. 6, 13, 35.)
- Mills A W (1958): "On the minimum audible angle". In: *J. Acoust. Soc. Am.* 30(4):237–246. (Cited on pp. 4, 55.)
- Minnaar P, Christensen F, Møller H, Oleson S K and Plogsties J (1999): "Audibility of all-pass components in binaural synthesis". In: *100th AES Convention, Preprint*. Munich, Germany. (Cited on p. 17.)
- Minnaar P, Oleson S K, Christensen F and Møller H (2001): "Localization with binaural recordings from artificial and human heads". In: *Journal of the Acoustic Engineering Society* 49(5):323–336. (Cited on pp. 2, 24.)
- Moldrzyk C (2002): "Ein neuartiger Kunstkopf zur Verifikation einer akustischen Entwurfsmethodik für Architekten". In: *22. Tonmeistertagung des Verbands Deutscher Tonmeister*. Hannover. (Cited on p. 10.)

- Moldrzyk C, Lentz T and Weinzierl S (2005): "Perzeptive Evaluation binauraler Auralisationen". In: *DAGA 2005: Plenarvorträge und Fachbeiträge der 31. Deutschen Jahrestagung Akustik*. München, Deutschland, pp. 545–546. (Cited on pp. 12, 24.)
- Møller H (1992): "Fundamentals of binaural technology". In: *Applied Acoustics* 36:171–218. (Cited on pp. 1, 6, 7, 8, 9, 13, 31.)
- Møller H, Hammershøi D, Jensen C B and Sørensen M F (1995a): "Design criteria for headphones". In: *J. Audio Eng. Soc.* 43(4):218–232. (Cited on p. 25.)
- Møller H, Hammershøi D, Jensen C B and Sørensen M F (1995b): "Transfer characteristics of headphones measured on human ears". In: *J. Audio Eng. Soc.* 43(4):203–217. (Cited on pp. 9, 15, 25, 26, 28, 29, 35, 48, 51.)
- Møller H, Hammershøi D, Jensen C B and Sørensen M F (1999): "Evaluation of artificial heads in listening tests". In: *J. Audio Eng. Soc.* 47(3):83–100. (Cited on p. 24.)
- Møller H, Jensen C B, Hammershøi D and Sørensen M F (1996a): "Using a typical human subject for binaural recording". In: *100th AES Convention*. Copenhagen, Denmark. (Cited on pp. 24, 25, 29.)
- Møller H, Sørensen M F, Hammershøi D and Jensen C B (1995c): "Head-related transfer functions of human subjects". In: *J. Audio Eng. Soc.* 43(5):300–321. (Cited on pp. 15, 25, 29, 79.)
- Møller H, Sørensen M F, Jensen C B and Hammershøi D (1996b): "Binaural technique: Do we need individual recordings?" In: *J. Audio Eng. Soc.* 44(6):451–469. (Cited on pp. 2, 23, 27, 29.)
- Moore B J C (1995): "Hearing". In: B J C Moore (Ed.) *Frequency analysis and masking*, Handbook of perception and cognition, 2nd ed. San Diego: Academic Press, pp. 161–205. (Cited on p. 59.)
- Möser M (2009): *Engineering acoustics*. 2nd ed. Berlin Heidelberg: Springer. (Cited on pp. 13, 36, 39.)
- Müller S (1999): *Digitale Signalverarbeitung für Lautsprecher*. Dissertation, RWTH Aachen, Germany. (Cited on pp. 20, 83.)
- Müller S (n.s. a): Building Exitation Signals (Monkey Forest Manual, http://www.fouraudio.com/data/MF/bex.pdf). Four Audio. (Cited on p. 36.)

- Müller S (n.s. b): Case study: Secondary Microphone Calibration by substitution method (Monkey Forest Manual, http://www.four-audio.com/data/MF/mic-cal.pdf). Four Audio. (Cited on p. 36.)
- Müller S and Massarani P (2001): "Transfer function measurement with Sweeps. Directors's cut including previously unreleased material and some corrections". In: *J. Audio Eng. Soc.* (Original release) 49(6):443–471. (Cited on pp. 36, 49.)
- Norcross S G, Bouchard M and Soulodre G A (2006): "Inverse Filtering design using a minimal phase target function from regularization". In: *121th AES Convention, Convention Paper 6929*. San Francisco, USA. (Cited on pp. 19, 20, 22, 30, 53, 54, 64, 80, 89.)
- Oppenheim A, Schafer R W and Buck J R (2004): *Zeitdiskrete Signalverarbeitung*. 2nd ed. Munich et. al: Pearson Education. (Cited on pp. 7, 11.)
- Paquier M and Koehl V (2010): "Audibility of headphone positioning variability". In: *128th AES Convention, Convention Paper 8147*. London, UK. (Cited on pp. 26, 27, 29, 51, 93.)
- Pralong D and Carlile S (1994): "Measuring human head-related transfer functions: A novel method for the construction and calibration of a miniature "in-ear" recording system". In: *J. Acoust. Soc. Am.* 95(6):3435–3444. (Cited on pp. 14, 35.)
- Pralong D and Carlile S (1996): "The role of individualized headphone calibration for the generation of high fidelity virtual auditory space". In: *J. Acoust. Soc. Am.* 100(6):3785–3793. (Cited on pp. 1, 2.)
- Rausch J (2008): Neuvermessung von HRTFs unter Berücksichtigung der Eingangsimpedanz des Ohrkanals. Master Thesis, Carl von Ossietzky Universität Oldenburg, Germany. (Cited on p. 35.)
- Rec ITU-R BS1116-1 (1997): "Methods for the subjective assement of small impairments in audio systems including multichannel sound systems". (Cited on pp. 64, 67, 72, 73, 85, 86.)
- Riederer K A J (1998): "Repeatability analysis of head-related transfer function measurements". In: *105th AES Convention*. San Francisco, USA. (Cited on pp. 13, 16, 30, 31.)
- Riederer K A J (2004a): "Part IIIa: Effect of microphone position changes on blocked cavum conchae head-related transfer functions". In: *18th Intern. Congress on Acoustics*. Kyoto, Japan, pp. 787–790. (Cited on pp. 15, 16, 30, 35.)

- Riederer K A J (2004b): "Part IVa: Effect of Cavum Conchae Blockage on Human Head-Related Transfer Functions". In: *18th Intern. Congress on Acoustics*. Kyoto, Japan, pp. 791–794. (Cited on pp. 15, 32, 48.)
- Rotter A (2010): Wahrnehmbarkeit klanglicher Unterschiede von Hochtonlautsprechern unterschiedlicher Wirkprinzipien. Magisterarbeit, Technische Universität Berlin, Germany. (Cited on p. 52.)
- Schärer Z (2008): Kompensation von Frequenzgängen im Kontext der Binauraltechnik. Magisterarbeit, Technische Universität Berlin, Germany. (Cited on pp. 1, 2, 11, 12, 13, 17, 18, 19, 20, 21, 22, 24, 26, 27, 29, 30, 48, 51, 53, 55, 59, 63, 64, 65, 66, 73, 82, 103, 105.)
- Schärer Z and Lindau A (2009): "Evaluation of Equalisation Methods for Binaural Signals". In: *126th AES Convention, Convention Paper*. Munich, Germany. (Cited on pp. 17, 54.)
- Schultz F (2011): *Elektroakustisch optimierte Wiedergabe binauraler Signale*. Master Thesis, Technische Universität Berlin, Germany. (Cited on pp. 52, 92.)
- Shaw E G A (1998): "Acoustical characteristics of the outer ear". In: M Crocker (Ed.) *Handbook of acoustics*, 1st ed. John Wiley & Sons, pp. 1093–1103. (Cited on p. 24.)
- Shaw E G A and Teranish R (1968): "Sound Pressure Generated in an External-Ear Replica and Real Human Ears by a Nearby Point Source". In: *Journal of the Acoustic Society of America* 44(2):240–249. (Cited on p. 24.)
- Slayney M (1998): "Auditory toolbox. Version 2". Technical report, Interval Research Corporation. (Cited on p. 59.)
- Strutt (Lord Rayleigh) J W (1907): "On our perception of sound direction". In: *Philosophical Magazine* 13:214–232. (Cited on p. 3.)
- Theile G (1986): "On the standardization of the frequency response of high-quality studio headphones". In: *J. Audio Eng. Soc.* 34(12):956–969. (Cited on p. 25.)
- Toole F E and Sayers B M (1965): "Lateralization judgements and the nature of binaural acoustic images". In: *J. Acoust. Soc. Am.* 37:319–324. (Cited on p. 4.)
- Voogdt U (2005): *Otoplastik. Die individuelle Otoplastik zur Hörgeräteversorgung und als persönlicher Gehörschutz im Lärm.* Fachwissenschaftliche Reihe der Akademie für Hörgeräte-Akustik Band 2, 3rd ed. Median Verlag. (Cited on pp. 32, 34, 45.)

- Vorländer M (2008): Auralization. Fundamentals of acoustics, moddeling, simulation, algorithmics and acoustic virtual reality. 1st ed. Berlin Heidelberg: Springer. (Cited on pp. 4, 5, 8.)
- Wenzel E, Arruda M, Kistler D and Wightman F (1993): "Localization using nonindividualized head related transfer functions". In: *J. Acoust. Soc. Am.* 94(1):111–123. (Cited on p. 27.)
- Wettschurek R (1970): "Über Unterschiedsschwellen beim Richtungshören in der Medianebene". In: *Gemeinschaftstagung für Akustik und Schwingungstechnik, Berlin*. Düsseldorf: VDI-Verlag, pp. 385–388. (Cited on p. 4.)
- Whitworth R H and Jeffress L A (1961): "Time versus intensity in the localization of tones". In: *J. Acoust. Soc. Am.* 33:925–929. (Cited on p. 4.)
- Wightman F and Kistler D (1989): "Headphone simulation of free field listening. I:Stimulus synthesis". In: *J. Acoust. Soc. Am.* 85(2):858–867. (Cited on pp. 14, 16, 35.)
- Zotter F (2009): *Analysis and synthesis of sound-radiation with spherical arrays*. Dissertation, University of music and performing arts, Graz. (Cited on pp. 10, 30.)

List of Figures

2.1.	Head-related spherical coordinate system	4
2.2.	Sound transmission through the external ear	7
2.3.	Recording and reproducing binaural signals	9
2.4.	Auralization using dynamic binaural synthesis	11
2.5.	Examples for microphone positioning from various studies	14
2.6.	Target bandpass used by Schärer (2008)	18
2.7.	Compensation results from Schärer (2008)	21
2.8.	HRTFs of 43 human subjects	24
2.9.	HPTFs of 40 human subjects and 3 headphones	26
2.10	.HPTFs measured on the left ear of the FABIAN HATS	27
3.1.	Anatomic sketch of the outer ear	32
3.2.	3D model of parts from a left outer ear	33
3.3.	Knowles FG-23329 frequency responses	38
3.4.	DPA 4060 frequency responses	41
3.5.	DPA 4060 microphones measured in 2006 and in 2011	42
3.6.	Prototype of left ear mould with flush-cast microphone	43
3.7.	Negative mould: Mesh Grid and 3D plot	44
3.8.	Crafting of the measurement instrument	46
3.9.	Measurements conducted on an artificial ear	47
3.10	PRECISE ear moulds on human ears	49
3.11	.Intra individual variance in HPTFs	50
3.12	.Inter individual variance in HPTFs	52
3.13	Non-indivdual and generic headphone compensation filter	55
3.14	Individual headphone compensation of subject 17	56
3.15	Generic headphone compensation of subject 17	57
3.16	Non-individual headphone compensation of subject 17	58
3.17	Auditory filterbank consisting of 40 ERB filters	59
3.18	Modeled auditory deviations from target function	60
4.1.	ABC/HR GUI	65
4.2.	Audio content	66
4.3.	Setup – listening test I	68
4.4.	Signal flow chart – listening test I	69

4.5.	Physical validation of electro acoustical setup and signal processing	71
4.6.	Attributes allowing the distinction between real and simulated soundfield –	
	listening test I	74
4.7.	Results: Mean differences grades and 95% confidence intervals across condi-	
	tions – listening test I	75
4.8.	Significant interactions – Listening test I	77
4.10	.Modeled auditory deviations of compensated HPTFs from target function	85
4.11	.Attributes allowing the distinction between real and simulated soundfield –	
	listening test II	86
4.12	Results: Mean differences grades and 95% confidence intervals across condi-	
	tions – listening test II	87
4.13	.Significant interaction – Listening test II	88

List of Tables

2.1.	Localization and lateralization blur	5
2.2.	Sound pressures at the external ear	8
2.3.	Attributes allowing the distinction between real and simulated sound field	12
2.4.	Influences on the quality of binaural simulation	30
3.1.	Measures of the ear canal	33
3.2.	Mean and percentile values of ear canal measures	34
3.3.	Miniature Microphones	35
3.4.	Measures used for negative mould	44
3.5.	Individual, generic and non-individual beta weights	54
3.6.	Modeled auditory deviations of compensated HPTFs from target functions	61
4.1.	Conditions of Schärer (2008) and present study – listening test I	63
4.2.	Parametric equalizers used for subwoofer and room equalization	72
4.3.	Main effects and significant interactions – Listening test I	76
4.4.	Conditions of listening test II	80
4.5.	Main effects and significant interactions – Listening test II	88
B.1.	A priori sample size estimation - Listening test I	18
B.2.	Post hoc power estimation - Listening test I	18
B.3.	Answers from Questionnaire – Listening test I	.19
B.4.	A priori sample size estimation - Listening test II	25
B.5.	Post hoc power estimation - Listening test II	25
B.6.	Answers from Questionnaire – Listening test II	.27

Appendix A

Specification of technical equipment

Knowles FG-23329





APPENDIX A. SPECIFICATION OF TECHNICAL EQUIPMENT

FG-23329-P07 SHEET 2.1 -40 dB -45 -50 -55 -60 -65 -70 -75 -80 2 3 4 5 6 789 2 3 4 5 6789 00 1000 0000 FREQUENCY IN HERTZ SENSITIVITY DEVICE CONFORMITY FREQUENCY MIN. NOM MAX. RANGE OF DEVIATION FROM IKHz -53.0 -53.0 -53.0 00| 1000 10000 -3.0 0.0 -3.0 +3.0 0.0 +3.0 -50.0 -56.0 NOTES: I. CASE CONNECTED TO NEGATIVE TERMINAL. 2. MICROPHONE TO BE FUNCTIONAL WITH 1.6 VDC SUPPLY. 3. TYPICAL SENSITIVITY TO HUMIDITY AT 1000Hz IS 0.06 dB/XRH. 4. SENSITIVITY AND NOISE VALUES INDICATED ON THIS SPECIFICATION ARE VALID AT 50% HUMIDITY. SENSITIVITY CHANGE ON REDUCING SUPPLY TO 0.9VDC OUTPUT IMPEDANCE OHMS CAPACITANCE ±50% PORT DC AMPLIFIER LOCATION SUPPLY CURRENT DRAIN "A" WEIGHTED NOISE (IKHz EQUIV. SPL) OUTPUT TO OUTPUT TO NEGATIVE POSITIVE MIN MAX. NOM 125 I.3V 50 µA MAX. 3 dB MAX. 30.0 dB MAX. 2800 4400 6800 l 3pF l 3pF CAPACITANCE PROBE POINTS

SENSITIVITY IN dB RELATIVE TO 1.0 VOLT/0.1 Pd (N/M²) FOR CONDITIONS SHOWN BELOW.



	Revision	C.O. #	Implementation Date	RELEASE LEVEL		REVISION
			10 10 07	Delessed		D
	A	MI0101572	11-21-05	Releasea		D
	WHEN TEST	LINITS ARE US	ED TO ESTABLISH INCOMING	INSPECTION ACCEPTANCE/REJECTION	DR. BY	DATE
WNOWI ES ELECTRONICS	ELIMINATION OF EQUIPMENT AND TEST NETHOD VARIATION				NNM	11-21-05
KNOWLES ELECTRONICS					CK. BY	DATE
ITASCA, ILLINOIS U.S.A.	TITLE:	ALC ROP	HONE/CABLE	FG-23329-P07	GJP	11-22-05
	· ·				APP. 81	DATE
		PERFORMAN	NCE SPECIFICATION	SHT 2.1	GJP	11-22-05

DPA 4060

Directional characteristics	Omnidirectional		
Principle of operation	Pressure		
Cartridge type	Pre-polarized condenser element with vertical diaphragm		
Frequency Range, $\pm 2 \text{ dB}$	Soft boost grid: 20 Hz – 20 kHz, 3 dB soft boost at 8 – 20 kHz.		
	High boost grid: 20 Hz – 20 kHz, 10 dB boost at 12 kHz.		
Sensitivity, nominal, $\pm 3 \text{ dB}$ at 1 kHz	20 mV/Pa; -34 dB re. 1 V/Pa		
Equivalent noise level A-weighted	Typ. 23 dB(A) re. 20 µPa (max. 26 dB(A))		
Equiv. noise level ITU-R BS.468-4	Typ. 35 dB (max. 38 dB)		
Dynamic range	Typ. 100 dB		
Max. SPL, peak before clipping	134 dB		
Microphone diameter	5.4 mm (0.21 in)		
Microphone length	12.7 mm (0.5 in)		
Technical Specification of DPA 4060			



Directional Characteristics of DPA 4060 (normalized)

Brüel & Kjær 4135

Calibration Chart for Brüel & Kjær **Condenser Microphone** Conditions of Tests: Anna в Cartridge Type 4135 Ξĸ Frequency: 250 Hz ALC: NO Polarization voltage: 200 V Nærum Denmark Serial No.800563..... Relative Humidity: 49 * Open Circuit Sensitivity at 1013 mbar, 23°C and 24 ΙĊ. Temperature -49.1 dB re. 1 V per Pa or 3.51 mV per Pa **Open Circuit Correction Factor:** Ko*) =+23.1..... dB 14. Cartridge Capacitance: Leekage Resistance tested at 52% relative humidity Summarized Specifications Outside Diameter: 0,25 inch (6,35 mm) without pro-tecting grid Frequency Response Characteristics: The upper curve is the open circuit free-field charac-teristic, valid for the Microphone Cartridge without protecting grid. Sound waves perpendicular to diaph-ragm (see Fig.). The lower curve is the open circuit pressure re-sponse recorded with electrostatic actuator. The rade curve is the open circuit random incidence response. Protection grid included. Grid Thread (coupler mounting): 0.25 inch (6.35 mm), 60 UNS2 Resonance Frequency: approx. 100 kHz Equivalent Air Volume at 1 atm.: less than 0.0005 cm² Temperature coefficient between -50 and + 60°C: Less than ± 0.01 dB/°C 4135 ((((() 淼 Ambient Pressure: Influence on sensitivity less than ---0,1 dB for + 10% pressure change *) Subtract the gain of the preamplifier (se from Ko to get the actual cirrection factor manual for the use of K). Relative Humidity: Influence less than 0,1 dB, in the absence of cendensation of this card) Limiting Sound Pressure: 185 dB (not tested at higher levels) 1 Pa = 1 N/m² = 10 dynes/cm² = 10 µbar BC 0036 Measuring Object -. -. . DA B C Lin. Zero Level: Date: Potentiometer: Sign.: +5 ... 1 18 T. T -T ×. dB +1-0. -1--5 -10 -15 -20 20 100000 100 1000 2000 5000 10000 20000 50000 500 50 200 3 Rectifer. ____mm/sec. Paper Speed:___ mm/sec. ___ Lower Lim. Freq .:... ____Hz Writing Speed:__

Genelec 8030a

8030A Data Sheet



STSTEM SPECIFICATIONS				
	8030A			
Lower cut-off frequency, -3 dB Upper cut-off frequency, -3 dB Free field frequency response	≤ 55 Hz ≥ 21 KHz 58 Hz – 20 kHz (± 2.0 dB)			
Maximum short term sine wave acoustic output on axis in half space, averaged from 100 Hz to 3 kHz	@ 1 m ≥ 100 dB SPL @ 0.5 m ≥ 106 dB SPL			
Maximum long term RMS acoustic output in same conditions with IEC-weighted noise (limited by driver unit protection circuit)	@ 1 m ≥ 97 dB SPL			
Maximum peak acoustic output per pair @ 1 m from the listening position with music material	≥ 108 dB			
Self generated noise level in free field @ 1 m on axis	≤ 10 dB (A-weighted)			
Harmonic distortion at 85 dB SPL @ 1 m on axis	Freq: 50100 Hz < 2 % > 100 Hz < 0.5 %			
Drivers	Bass 130 mm (5") cone Treble 19 mm $(^3/_4")$ metal dome Both drivers are magnetically shielded			
Weight	5.6 kg (12.3 lb)			
Speaker dimensions	Height 299 mm (11 ¹⁵ / ₁₅) (including Iso-Pod table stand) Height 285 mm (11 ¹ / ₁) (without Iso-Pod table stand) Width 189 mm (7 ⁷ / ₁₅)			
	Depth 178 mm (7").			



and "bass roll-off" controls on the free field response.

AMPLIFIER SECTION	
	8030A
Bass amplifier output power Treble amplifier output power	Short term 40 W (8 Ohm load) Short term 40 W (8 Ohm load) Long term output power is limited by driver unit protection circuitry.
Amplifier system distortion at nominal output	THD ≤0.05 % SMPTE-IM ≤0.05 % CCIF-IM ≤0.05 % DIM 100 ≤0.05 %
Signal to Noise ratio, referred to full output	Bass ≥100 dB Treble ≥100 dB
Mains voltage: Voltage operating range:	100, 120, 220 or 230 V according to region. ±10 %
Power consumption	Idle 10 VA Full output 80 VA

CROSSOVER SECTION		
	8030A	
Input connector: XLR female, balanced 10 kOhm	pin 1 gnd, pin 2 +, pin 3 -	
Output connector: XLR male, balanced 100 kOhm	pin 1 gnd pin 2 +, pin 3 -	
Input level for 100 dB SPL output @ 1m	-6 dBu at volume control max	
Volume control range	-80 dB relative to max output Output signal level is 0 dB relative to input signal level but adjustable by volume control	
Crossover frequency	3.0 kHz	
Treble tilt control operating range	0 to -2 dB @ 15 kHz	
Bass roll-off control	-6 dB step @ 85 Hz (to be used in conjunction with a 7050A subwoofer)	
Bass tilt control	0 to -6 dB @ 100 Hz in 2 dB steps	
	The 'CAL' position is with all tone controls set to 'off' and input sensitivity control to maximum and corresponds to a maximally flat free field response.	

Genelec Document BBA0036001a Copyright Genelec Oy 4 2007 All data subject to change without prior notice



International enquiries: Genelec, Olvitie 5 FIN 74100, lisalmi, Finland Phone +358 17 83881 Fax +358 17 812 267 Email genelec@genelec com

Fax +1 508 652 0909

Email genelec usa@gen

In the U S please contact: Genelec, Inc , 7 Tech Circle Natick, MA 01760, U S A Phone +1 508 652 0900 In China please contact: Beijing Genelec Audio Co Ltd Jianwai SOHO, Tower 12, Room 2306 39 East 3rd Ring Road Chaoyang District, Beijing 100022, China Phone +86 10 5869 7915 Fax +86 10 5869 7914

www.genelec.com In Sweden please contact: Genelec Sverige Box 5521, S 141 05 Huddinge Phone +46 8 449 7070 Fax +46 8 708 7071 Email info@genelec con







8030Avlews 13.04.04/MKu + 8.11.06/SDM

Polhemus Fastrak©

Degrees of freedom	6 (X, Y, Z, azimuth, elevation, roll)		
Number of sensors	4		
Update rate	120 Hz divided by number of sensors used		
Static accuracy position	0.03 inch RMS		
Static accuracy orientation	0.15° RMS		
Latency	4 ms		
Resolution position at 30 cm range	Resolution position per inch of source and sensor separation		
Range from standard TX2 source	Up to 1.52 m		
Extended range source	Up to 4.6 m		
Interface	RS-232		
Technical Creation of Delhamus Festual			

Technical Specification of Polhemus Fastrak®

Genelec 1031a

1031A SYSTEM SPECIFICATIONS

Lower cut-off frequency, -3 dB:< 47 Hz

		8 Ohm load:			
Upper cut-off frequency,	-3 dB: <u>></u> 22 kHz			120 W	
Free field frequency response		Treble amplifier output power with an			
or system: 40 Hz -	22 KHZ (± 2 UD)	o onin loda.		120 W	
Maximum short term sine wave acoustic output on axis in half space, averaged from 100 Hz to 3 kHz:		Long term output power is limited by drive unit protection circuitry.			
@ 1m	> 110 dB SPL				
@ 0.5m	≥ 116 dB SPL	Slew rate :		80 V/µs	
Maximum long term RMS output in same conditions	acoustic with IEC	Amplifier system distortion at nominal output:			
weighted noise (limited by	y driver unit pro-		THD	≤ 0.05%	
@ 1m	< 101 dB SPI		SMPTE-IM	<u><</u> 0.05%	
@ 0.5m	> 107 dB SPI		CCIF-IM	<u><</u> 0.05%	
w 0 Jili	<u>2</u> 107 08 5FL		DIM 100	<u><</u> 0.05%	
Maximum peak acoustic of per pair on top of console	output ə,	Signal to Noise ratio, referred to full output:			
with music material:	<u>></u> 120 dB		Bass Treble	≥ 100 dB ≥ 100 dB	
Self generated noise level	l in		/		
free field @ 1m on axis:	<u><</u> 10 dB	Mains voltage:	100/200 V or	115/230 V	
	(A-weighted)	Voltage operatio	a ranae at		
Llaurania distantiana at 00		230V setting	207 - 253 V	(+ 10%)	
Harmonic distortion at 90	ab SPL @ 1m on	115V setting:	104 - 126 V	$(\pm 10\%)$	
Freq: 50 100 Hz	< 1%	riot ootang.	101 1201	(10,0)	
> 100 Hz	< 0.5%	Power consump	tion:		
100112	100,0		Idle	30 W	
Drivers: Bass 210 mm	(8") cone		Full output	160 W	
Treble 25 mm	(1") metal dome				
Both drivers are magnetic	ally shielded				
Weight: 12,7 kg	g (28 lb)				
Dimensions:					

AMPLIFIER

Bass amplifier output power with an

maximum.

SECTION

CROSSOVER SECTION

Input connector: XLR female pin 1 gnd pin 2 +
pin 3 -
Input impedance: 10 kOhm balanced
Input level for 100 dB SPL output @ 1m: variable from +6 to -6 dBu
Input level for maximum short term output of 110 dB SPL @ 1m: variable from +16 to +4 dBu
Subsonic filter below 45 Hz : 18 dB/octave
Ultrasonic filter above 25 kHz: 12 dB/octave
Crossover frequency, Bass/Treble:2.2 kHz
Crossover acoustical slopes: 24 dB/octave
Treble tilt control operating range in 2 dB steps: from +2 to -4 dB & MUTE
Bass roll-off control operating range in 2 dB steps: from 0 to -8 dB @ 40 Hz
Bass tilt control operating range in 2 dB steps: from 0 to -6 dB & MUTE
The 'CAL' position is with all tone controls set to 'off' and input sensitivity control to

Height 395 mm (15 ⁹/₁₆") Width 250 mm (9 ⁷/₈") Depth 290 mm (11 7/16")

Genelec Oy, Olvitie 5 FIN - 74100 IISALMI, FINLAND Phone: +358 - 17 - 813311 Telefax: +358 - 17 - 812267 Email: genelec@genelec.com Web: http://www.genelec.com



Data Sheet No. 1031-0107-6 COPYRIGHT GENELEC OY 1997 All data subject to change wi hout prior notice



Knowles FG-23329 power supply

Circuit diagram. Switch position: Knowles operation mode.



Picture. Switch position: Knowles operation mode.

Appendix B

Perceptual Evaluation

Statistical data - Listening test I

F tests - ANOVA: Fixed effects, omnibus, one-way					
Anlysis:	A priori: Compute required s	sample size			
Input:	Effect size f α err prob Power (1- β err prob) Number of groups	= 0.1291 = 0.05 = 0.8 = 2			
Output:	Noncentrality parameter λ Critical F Numerator df Denominator df Total sample size Actual power	= 7.9000679 = 3.8612355 = 1 = 472 = 474 = 0.8009507			

Table B.1. A priori sample size estimation - Listening test I.(Obtained with G*Power 3 (Faul et al., 2007))

F tests - ANOVA: Fixed effects, omnibus, one-way

Anlysis:	Post hoc: Compute achieved	power
Input:	Effect size f	= 0.1291
	α err prob	= 0.05
	Total sample size	= 600
	Number of groups	= 2
Output:	Noncentrality parameter λ	= 10.0000860
	Critical F	= 3.8570560
	Numerator df	= 1
	Denominator df	= 598
	Actual power	=0.8843946
	Actual power	=0.8843946

Table B.2. Post hoc power estimation - Listening test I.(Obtained with G*Power 3 (Faul et al., 2007))

Subject	Attributes that were used to distinguish between real sound field and simulation
1	Rauschen: Helligkeit, zusätzlich: tieffrequentes Rauschen, Dumpf belegt
2	Klangfarbe: weniger Hall (genauer: brilliant), komplex verfärbt, Lokalisation: näher. Klangfarbe: weniger transparent (Gitarrenstimulus). Lokalisation: etwas instabiler, insbesondere ohne Kopfbewegung, weiter rechts. Die Spezifizierung bezieht sich jeweils auf den kontaminierten Reiz.
3	Entfernung: Simulation ist näher als Lautsprecher, Bässe/tiefe Frequenzen (Git. Saiten) klingen schwächer ("Grundton fehlt")
4	Klangfarbe, Ortung, Räumlichkeit, klangfarbliche Veränderung bei Kopfdrehung
5	Räumliche Abgrenzung der Quelle/Räumliche Schärfe. Simulation dumpfer bzw. mehr untere Mitten/Bässe. Subwoofer war als solcher zu erkennen
6	Klangfarbe, Lautstärke
7	Leider nur Klangfarbe, spezifischer: Spektralverteilung, Höhenanteil
8	Breite, Farbe, Filterung
9	Rauschen: Tonhöhe, Nähe/Ferne, Quellbreite. Musik: Nähe/Ferne, Quellbreite
10	Kopfhörer war meist druckvoller/gefühlt lauter, manchmal ein bisschen Blechern, genauer als Lautsprecher
11	Lautstärke: Die Simulation war für mich lauter, Klangfarbe. Räumlichkeit: Simulation war etwas luftiger
12	Klangfarbe, Lautstärke
13	Lautsärke, Klangfarbe, Intensität
14	Die Simulation war oft in ihrer Brillianz weniger ausgeprägt (also Höhen), Die Simu-
	lation wirkte oft grobkörniger (Samplingfrequenz/Bit-Rate?)
15	Einschwingverhalten, Lautstärke, besonders im Bassbereich. Klangfarbe Bass
16	Rauschen: extreme Färbung 1-2kHz. Bei Simulation Artefakte in der Mitte des Stimu- lus. Verschmiert. Gitarre: spektrale Probleme zu viel Bass, zu nasal
17	Richtung, Klangfarbe, Räumlichkeit
18	Klangfarbe (nur bei Rauschen). Bei Gitarre: Lokalisation/Nähe (Punktförmige Fokussierung vs. Räumlich verschmierte Schallquelle). Empfindung von Transien- ten
19	Höhenabfall, Rumpeln im Bass bei 2-3 Rauschstimuli, Lautstärke
20	Klangfarbe. Simulation: dumpfer (Rauschen), In den mitten überbetont und dumpfer (Gitarre)
21	Klangfarbe: Original klingt brillianter -> Höhenreicher. Räumlichkeit: Original klingt breiter + luftig
22	Klangfarbe (Simulation meistens dumpfer). Subwoofereinsatz (?) vor allem bei weiten Kopfbewegungen
23	Klangfarbe, Lokalisation, Teilweise Bässe nicht externalisiert
24	Lautstärke. Lokalisation eher von links. Unschärfe (weniger präsent)
25	Rauschen: Grummen tiefer Frequenzen. Gitarre: Picking Sounds und Gitarrenkorpus- Schlag

 Table B.3. Answers from Questionnaire – Listening test I.



Histogram for each condition – Listening test I, Read = not normal distributed.

Measure MEASURE_1							
						Eps lon(a)	
		Approx Ch -			Greenhouse-		
W th n Subjects Effect	Mauchly's W	Square	df	Sg	Ge sser	Huynh-Feldt	Lower-bound
content	1 000	000 0	0	1	1 000	1 000	1 000
f Iter	0 984	0 364	2	0 834	0 985	1 000	0 500
reproduct on	1 000	000 0	0	1	1 000	1 000	1 000
phase	1 000	000 0	0	1	1 000	1 000	1 000
content * f Iter	0 748	6 676	2	0 035	662 0	0 847	0 500
content * reproduct on	1 000	000 0	0	L	1 000	1 000	1 000
f Iter * reproduct on	0 992	0 192	2	0 908	0 992	1 000	0 500
content * f Iter * reproduct on	0 913	2 082	2	0 353	0 920	0 933	0 500
content * phase	1 000	000 0	0	1	1 000	1 000	1 000
f lter * phase	206 0	2 255	2	0 324	0 915	0 986	0 500
content * f Iter * phase	0 928	1 716	2	0 424	0 933	1 000	0 200
reproduct on * phase	1 000	000 0	0	L I	1 000	1 000	1 000
content * reproduct on * phase	1 000	000 0	0	1	1 000	1 000	1 000
f Iter * reproduct on * phase	606 0	2 197	2	0 333	0 917	0 989	0 500
content * f Iter * reproduct on * phase	0 983	0 396	2	0 821	0 983	1 000	0 500
Tests the null hypothes s that the error	or covar ance ma	itr x of the orthor	normal zed trans	formed depende	ent var ables s p	roport onal to an	i dent ty matr x
a May be used to adjust the degrees Effects table	of freedom for t	he averaged tes	ts of s gn f canc	e Corrected test	ts are d splayed	n the Tests of W	/ th n-Subjects
W th n Subjects Des gn							
content+f Iter+reproduct on+phase+c ntent*f Iter*phase+reproduct on*phas	ontent*f lter+con e+content*repro	tent*reproduct or duct on*phase+f	n+f lter*reprodu ter*reproduct o	ct on+content*f It n*phase+conter	<pre>ter*reproduct on+ it*f Iter*reproduct</pre>	+content*phase+ t on*phase	f Iter*phase+∞

Mauchly's Test of Sphericity(b)

SPSS output: Mauchly's test for sphericity - Listening test I.

121

			Tests	of Within-Subjec	ts Effects				
Measure MEA	SURE_1								
		Type Sum of					Partal Eta	Noncent	Observed
Source		Squares	df	Mean Square	ш	Sg	Squared	Parameter	Power(a)
content	Spher c ty Assumed	159 095	1	159 095	45 055	000 0	0 652	45 055	1 000
	Greenhouse-Ge sser	159 095	1 000	159 095	45 055	0 000 0	0 652	45 055	1 000
	Huynh-Feldt	159 095	1 000	159 095	45 055	0 000	0 652	45 055	1 000
	Lower-bound	159 095	1 000	159 095	45 055	0 000	0 652	45 055	1 000
f Iter	Spher c ty Assumed	43 639	2	21 820	19 848	000 0	0 453	39 697	1 000
	Greenhouse-Ge sser	43 639	1 969	22 162	19 848	0 000	0 453	39 083	1 000
	Huynh-Feldt	43 639	2 000	21 820	19 848	0 000	0 453	39 697	1 000
	Lower-bound	43 639	1 000	43 639	19 848	0 000	0 453	19 848	066 0
reproduct on	Spher c ty Assumed	2 503	1	2 503	2 008	0 169	0 077	2 008	0 275
	Greenhouse-Ge sser	2 503	1 000	2 503	2 008	0 169	0 077	2 008	0 275
	Huynh-Feldt	2 503	1 000	2 503	2 008	0 169	0 077	2 008	0 275
	Lower-bound	2 503	1 000	2 503	2 008	0 169	0 077	2 008	0 275
phase	Spher c ty Assumed	0 756	1	0 756	2 146	0 156	0 082	2 146	0 290
	Greenhouse-Ge sser	0 756	1 000	0 756	2 146	0 156	0 082	2 146	0 290
	Huynh-Feldt	0 756	1 000	0 756	2 146	0 156	0 082	2 146	0 290
	Lower-bound	0 756	1 000	0 756	2 146	0 156	0 082	2 146	0 290
reproduct on *	Spher c ty Assumed	4 765	1	4 765	096 6	0 004	0 293	096 6	0 857
phase	Greenhouse-Ge sser	4 765	1 000	4 765	996 6	0 004	0 293	096 6	0 857
	Huynh-Feldt	4 765	1 000	4 765	996 6	0 004	0 293	096 6	0 857
	Lower-bound	4 765	1 000	4 765	9 960	0 004	0 293	996 0	0 857
f Iter	Spher c ty Assumed	1 877	2	0 939	4 536	0 016	0 159	9 072	0 746
*reproduct on *	Greenhouse-Ge sser	1 877	1 833	1 024	4 536	0 019	0 159	8 315	0 718
pnase	Huynh-Feldt	1 877	1 977	0 950	4 536	0 016	0 159	8 968	0 743
	Lower-bound	1 877	1 000	1 877	4 536	0 044	0 159	4 536	0 534
a Computed u	s ng alpha = 05								

SPSS output: Within subject effects (shortened) – Listening test I.

122

Estimated Marginal Means

1. Grand Mean

Measure: MEASURE_1

		95% Confide	nce Interval
Mean	Std. Error	Lower Bound	Upper Bound
2,21E 007	,000	1,84E 006	1,40E 006

2. content

Estimates

Measure: MEASURE 1

			95% Confide	ence Interval
content	Mean	Std. Error	Lower Bound	Upper Bound
1	,515	,077	,673	,357
2	,515	,077	,357	,673

Pairwise filterarisons

Measure: MEASURE 1

		Mean			95% Confidence Inte	erval for Difference(a)
(I) content	(J) content	Difference (I J)	Std. Error	Sig	Lower Bound	Upper Bound
1	2	1,030(*)	,153	,000	1,347	,713
2	1	1,030(*)	,153	,000	,713	1,347

Based on estimated marginal means

* The mean difference is significant at the ,05 level.

a Adjustment for multiple filterarisons: Bonferroni.

3. filter

Estimates Measure: MEASURE 1

mousare. merk				
			95% Confide	ence Interval
filter	Mean	Std. Error	Lower Bound	Upper Bound
1	,366	,063	,236	,496
2	,090	,062	,217	,038
3	,276	,057	,393	,159

Pairwise filterarisons Measure: MEASURE 1

		Mean	· · · · · · · · · · · · · · · · · · ·		95% Confidence Int/	erval for Difference(a)
(I) content	(J) content	Difference (I J)	Std. Error	Sig	Lower Bound	Upper Bound
1	2	,455(*)	,111	,001	,170	,741
'	3	,642(*)	,103	,000	,377	,907
2	. 1'	,455(*)	,111	,001	,741	,170
L'	3	,187	,100	,225	,071	,445
3	, <u>1</u> ′	,642(*)	,103	,000	,907	,377
1	2	. ,187	,100	,225	,445	,071

Based on estimated marginal means

* The mean difference is significant at the ,05 level.

a Adjustment for multiple filterarisons: Bonferroni.

SPSS output: Pairwise comparisons part I (shortened) - Listening test I.

4. reproduction

Estimates Measure: MEASURE 1

model of mEric				
			95% Confide	ence Interval
reproduction	Mean	Std. Error	Lower Bound	Upper Bound
1	,065	,046	,159	,029
2	,065	,046	,029	,159

Pairwise filterarisons

Measure: MEASURE 1

		Mean			95% Confidence Inte	erval for Difference(a)
(I) content	(J) content	Difference (I J)	Std. Error	Sig	Lower Bound	Upper Bound
1	2	,129	,091	,169	,317	,059
2	1	,129	,091	,169	,059	,317

Based on estimated marginal means

a Adjustment for multiple filterarisons: Bonferroni.

5. phase

Estimates

Measure. MEAC				
			95% Confide	ence Interval
phase	Mean	Std. Error	Lower Bound	Upper Bound
1	,035	,024	,086	,015
2	,035	,024	,015	,086

Pairwise filterarisons

Measure: MEASURE 1

		Mean			95% Confidence Inte	erval for Difference(a)
(I) content	(J) content	Difference (I J)	Std. Error	Sig	Lower Bound	Upper Bound
1	2	,071	,048	,156	,171	,029
2	1	,071	,048	,156	,029	,171

Based on estimated marginal means

a Adjustment for multiple filterarisons: Bonferroni.

13. reproduction * phase

Measure: MEAS	SURE 1	-			
				95% Con	fidence Interval
reproduction	phase	Mean	Std. Error	Lower Bound	Upper Bound
1	1	,189	,066	,326	,052
	2	,060	,046	,035	,155
2	1	,118	,063	,011	,248
	2	,011	,058	,109	,131

15. filter * reproduction *	phase
-----------------------------	-------

Measure: MEASURE 1

					95% Confid	ence Interval
filter	reproduction	phase	Mean	Std. Error	Lower Bound	Upper Bound
1	1	1	,340	,095	,143	,537
		2	,384	,089	,202	,567
	2	1	,373	,116	,134	,611
		2	,366	,082	,197	,536
2	1	1	,449	,144	,747	,151
		2	,004	,127	,257	,266
	2	1	,106	,098	,096	,309
		2	,020	,091	,208	,168
3	1	1	,459	,089	,642	,276
		2	,209	,085	,384	,033
	2	1	,124	,103	,337	,089
		2	,314	,086	,490	,137

SPSS output: Pairwise comparisons part II (shortened) - Listening test I.

Statistical data - Listening test II

F tests - A	NOVA: Fixed effects, omnibus	, one-way
Anlysis:	A priori: Compute required	sample size
Input:	Effect size f α err prob Power (1- β err prob) Number of groups	= 0.1291 = 0.05 = 0.8 = 2
Output:	Noncentrality parameter λ Critical F Numerator df Denominator df Total sample size Actual power	= 7.9000679 = 3.8612355 = 1 = 472 = 474 =0.8009507

Table B.4. A priori sample size estimation - Listening test II.
(Obtained with G*Power 3 (Faul et al., 2007))

F tests - ANOVA: Fixe	d effects,	omnibus,	one-way
------------------------------	------------	----------	---------

Anlysis:	Post hoc: Compute achieved	power
Input:	Effect size f α err prob Total sample size Number of groups	= 0.1346 = 0.05 = 648 = 2
Output:	Noncentrality parameter λ Critical F Numerator df Denominator df Actual power	= 11.7399197 = 3.8570560 = 1 = 646 =0.9280327

Table B.5. Post hoc power estimation - Listening test II.
(Obtained with G*Power 3 (Faul et al., 2007))

Subject	Attributes that were used to distinguish between real sound field and simulation
1	Klingelartefartke, Dumpfheit(stark), Dumpfheit (schwach)sonst eigentlich nichts
2	Klangfarbe: Simulation dumpfer teilweise klingeln. Lokalisation: Bei Simulation eher
	Richtung TMT
3	Simulation war etwas "direkter", hatte weniger "Tiefe". Schlechte Klangfarbe. Lokalisa-
	tion alles gleich gut. Ein schlechtes Verfahren hat bei Rauschen geklingelt
4	Kammfiltereffekte während Kopfbewegung. Verzerrte Höhen sowohl bei Schlagzeug
	(einzelne Beckenschläge, Glocke) als auch bei Rauschen. Bei Ausrichtung des Kopfes
	schien das Signal nur noch auf einem Ohr hörbar, hauptsächlich bei Schlagzeugsignal.
5	Klangfarbe (Höhen) Dynamik
6	Klang, Bandbreite (teilweise Pfeifen bei Simulation). Gefühl von Ausdehnung der
	Quelle.
7	Der Ton war mehr von links. Der Ton klang bei der Simulation dumpfer. Das drehen
	des Kopfes hat den links rechts-Unterschied erhöht.
8	Klangfarbe, speziell bei Rauschen. Transienten beim Schlagzeug waren bei der Simu-
	lation dumpfer und manchmal verriselt, unnatürlich
9	Hochfrequenzanteil war bei Rauschen unangenehm. Eine Art Nachhall bei
	Schlagzeugwiedergabe (Snare Schlag) war nicht angenehm. Tieffrequenzanteil beim
	Schlagzeug war verzerrt
10	Klangfarbe, Lokalisaition. Schlagzeug: beste Qualität hauptsächlich, HF nicht ganz
	außer Kopf lokalisiert. Schlechte Qualität grausam, klangfarblich dumpf verwaschen.
	Rauschen: leichte Klingelartefakte bei bestimmten Verfahren, Quelllokalisation plausi-
	bel, Klangfarbe extrem dumpf bei schlechten Verfahren
11	Klangfarbe, v.a. Tiefpassfilterung. Phasenschmierereien Kammfilter
12	Die Simulation habe ich meistens dumpfer empfunden
13	Klangvolumen, Klangrichtung, beim Rauschen die Tonhöhe. Beim Schlagzeug die
	Simulation z.T. dumpfer und leiser
14	Klangfarbe, Simulation schien meist eher links abzulaufen, schwer zu sagen, die
	Lokalisation stimmte noch gut, aber es schien rechts irgendwie die Fülle zu fehlen im
	Vergleich zu links im selben Sample. Manchmal bei schlecht bewerteten Beispielen
	etwas, das wie stärkere Phasenartefakte klang
15	Allgemein: Simulation leicht links neben dem Lautsprecher. Schlagzeug: Becken war
	bei Simulation höher, klang generell metallischer, nicht so satt, künstlicher. Rauschen:
	höhrer/hellere Klangfarbe gefühlt noch störender
16	Lokalisation weiter vorne und unten. Klangfarbe: Simulation war dumpfer, artefakte
	beim Drehen des Kopfes

Subject	Attributes that were used to distinguish between real sound field and simulation
17	Rauschen: Simulation hat sich generell stark unterschieden, hohe Töne waren über-
	präsent. Schlagzeug war wesentlich dichter dran. Simulation war ein Stück weiter
	hinten als das Reale
18	Rauschen: Höhenverlust, Klingeln. Schlagzeug: Lokalisation (Original war viel
	fokussierter vorne, Sim. War teilweise "ausgebreitet" zwischen mir und dem LS,
	Höhenverlust
19	Das Rauschen war in der Simulation generell dumpfer. Das Schlagzeug klang in der
	Simulation weniger räumlich.
20	gefühlt Drums 1-2 Ls Breiten-Versatz nach links. Merkmale: spektrale Verteilung -
	Klangfarben tiefen
21	Leider nur Klangfarbe: Referenz ist deutlich heller, im Vergleich klingen alle Simula-
	tionen "gedeckelt", dumpfer und irgendwie kaputt
22	Räumlicher Eindruck. Bei der Simulation höre ich Reflexionen von der Seite und von
	hinten. Klangfarbe: Simulation hat weniger höhen, klingt dumpfer.
23	Klangausbreitung im Raum (ohne Bewegung des Kopfes). Klangfarbe
24	Klangfarbe: höhen haben gefehlt, hat sich gefiltert angehört
25	Klangfarbe (Höhen) Dynamik
26	Klangfarbe: Helligkeit, Schärfe oder gar klingeln, Lokalisation: Distanz (ggf. Tendenz
	zur Internalisierung), Schärfe. Manche Simulationen ein Ohr besser -> wird dadurch
	sofort unplausibel, da es wie "kopfhörer-Stereofonie" klingt.
27	Helligkeit, Schärfe, Transparenz

Table B.6 – continued from previous page

Table B.6. Answers from Questionnaire – Listening test II.



Histogram for each condition – Listening test II. Read = not normal distributed.

Mauchly's Test of Sphericity(b)

Measure MEASURE_1

						Eps lon(a)	
		Approx Ch -			Greenhouse-		
W th n Subjects Effect	Mauchly's W	Square	df	Sg	Ge sser	Huynh-Feldt	Lower-bound
content	1 000	000 0	0	1	1 000	1 000	1 000
f Iter	1 000	000 0	0	1	1 000	1 000	1 000
phase	1 000	000 0	0	1	1 000	1 000	1 000
regular zat on	0 828	4 7 19	2	0 094	0 853	206 0	0 500
content * f Iter	1 000	000 0	0	1	1 000	1 000	1 000
content * phase	1 000	000 0	0	1	1 000	1 000	1 000
f lter * phase	1 000	000 0	0	1	1 000	1 000	1 000
content * f Iter * phase	1 000	000 0	0	1	1 000	1 000	1 000
content * regular zat on	0 861	3 728	2	0 155	0 878	286 0	0 200
f Iter * regular zat on	0 836	4 487	2	0 106	0 859	0 914	0 200
content * f Iter * regular zat on	0 993	0 169	2	0 919	0 993	1 000	0 200
phase * regular zat on	0 944	1 429	2	0 489	0 947	1 000	0 200
content * phase * regular zat on	0 976	0 610	2	0 737	0 976	1 000	0 200
f Iter * phase * regular zat on	0 974	0 658	2	0 720	0 975	1 000	0 500
content * f Iter * phase * regular zat on	0 961	0 985	2	0 611	0 963	1 000	0 500
Tests the null hypothes s that the	he error covar an	ce matr x of the	orthonormal zed	transformed de	pendent var able	s s proportona	to an dent ty
matrx							
a May be used to adjust the de	grees of freedor	n for the averag	ed tests of s gn i	f cance Correcte	ed tests are d spl	ayed n the Test	s of W th n-
b Des on intercent							
W th n Subjects Des gn							
content+f Iter+phase+regular z content*f Iter*regular zat on+ph	at on+content*f It ase*regular zat c	er+content*phas	se+f Iter*phase+	content*f Iter*ph: +f Iter*phase*reg	ase+content*reg	ular zat on+f lter* ent*f lter*phase*	'regular zat on+ 'regular zat on

SPSS output: Mauchly's test for sphericity - Listening test II.

			Tests	s of Within-Subjec	ts Effects				
Measure ME/	ASURE_1								
		Type Sum of					Partal Eta	Noncent	Observed
Source		Squares	đ	Mean Square	Ľ	Sg	Squared	Parameter	Power(a)
content	Spher c ty Assumed	72 611	+	72 611	55 017	000 0	0 679	55 017	1 000
	Greenhouse-Ge sser	72 611	1 000	72 611	55 017	0 000 0	0 679	55 017	1 000
	Huynh-Feldt	72 611	1 000	72 611	55 017	0 000 0	0 679	55 017	1 000
	Lower-bound	72 611	1 000	72 611	55 017	0 000	0 679	55 017	1 000
flter	Spher c ty Assumed	203 532	+	203 532	110 765	000 0	0 810	110 765	1 000
	Greenhouse-Ge sser	203 532	1 000	203 532	110 765	0 000 0	0 810	110 765	1 000
	Huynh-Feldt	203 532	1 000	203 532	110 765	0 000 0	0 810	110 765	1 000
	Lower-bound	203 532	1 000	203 532	110 765	0 000 0	0 810	110 765	1 000
phase	Spher c ty Assumed	000 0	1	000 0	000 0	0 983	0000	000 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
	Greenhouse-Ge sser	0 000	1 000	0 000 0	0 000	0 983	0000	000 0	0 050
	Huynh-Feldt	0 000	1 000	0 000 0	0 0 0 0	0 983	0000	000 0	0 050
	Lower-bound	0 000	1 000	0 000	0 0 0 0	0 983	0000	000 0	0 050
regular zat on	Spher c ty Assumed	0 718	2	0 359	0 832	0 441	0 031	1 665	0 185
	Greenhouse-Ge sser	0 718	1 706	0 421	0 832	0 425	0 031	1 421	0 173
	Huynh-Feldt	0 718	1 814	0 396	0 832	0 431	0 031	1 510	0 178
	Lower-bound	0 718	1 000	0 718	0 832	0 370	0 031	0 832	0 142
flter *	Spher c ty Assumed	6 437	2	4 718	10 496	000 0	0 288	20 993	0 984
regular zat on	Greenhouse-Ge sser	9 437	1 718	5 493	10 496	0000	0 288	18 030	0 971
	Huynh-Feldt	9 437	1 828	5 163	10 496	0 000 0	0 288	19 184	0 977
	Lower-bound	9 437	1 000	9 437	10 496	0 003	0 288	10 496	0 877
a Computed t	s ng alpha = 05								

SPSS output: Within subject effects (shortened) – Listening test II.

Tests of Between-Subjects Effects

Econtentated Marginal Means

Measure: MEASURE_1	
--------------------	--

		95% Confidence Interval		
Mean	Std. Err.	Lower Bound	Upper Bound	
0.000	0.000	1,53E 006	1,60E 006	

2. content

Econtentates

Measure: MEASURE_1							
95% Confidence Interval							
content	Mean	Std. Error	Lower Bound	Upper Bound			
1	,335	,045	,428	,242			
2	,335	,045	,242	,428			

Measure: MEASURE 1							
		Mean				Partial Eta	
(I) content	(J) content	Difference (I J)	Std. Error	Sig.	Sig.	Squared	
1	2	,669(*)	,090	,000	,855	,484	
2	1	,669(*)	,090	,000	,484	,855	

*. The mean difference is significant at the ,05 level.

a. Adjustment for multiple comparisons: Bonferroni.

3. filter

Econtentates

Measure: MEASURE_1								
			95% Confide	ence Interval				
filter	Mean	Std. Error	Lower Bound	Upper Bound				
1	0.560	0.053	0.670	0.451				
2	0.560	0.053	0.451	0.670				

Pairwise Comparisons

Measure: MEASURE_1							
		Mean			Differe	nce(a)	
(I) filter	(J) filter	Difference (I J)	Std. Error	Sig.(a)	Lower Bound	Upper Bound	
1	2	1,121(*)	,107	,000	1,340	,902	
2	1	1,121(*)	,107	,000	, <mark>902</mark>	1,340	

*. The mean difference is significant at the ,05 level.

a. Adjustment for multiple comparisons: Bonferroni.

SPSS output: Pairwise comparisons part I (shortened) - Listening test II.

4. phase

		Econtentates							
Measure: MEAS	Neasure: MEASURE 1								
			95% Confide	ence Interval					
phase	Mean	Std. Error	Lower Bound	Upper Bound					
1	0.001	0.026	0.053	0.05					
2	0.001	0.026	0.054	0.05					

Pairwise Comparisons

Measure: MEASURE 1								
		Mean			Differe	nce(a)		
(I) phase	(J) phase	Difference (I J)	Std. Error	Sig.(a)	Lower Bound	Upper Bound		
	1	2,001	,052	,983	,106	,108		
	2	1,001	,052	,983	,108	,106		

a. Adjustment for multiple comparisons: Bonferroni.

5. regularization

Measure: MEASURE 1

Econtentates

			95% Confidence Interval					
regularization	Mean	Std. Error	Lower Bound	Upper Bound				
1	0.001	0.029	0.060	0.058				
2	0.040	0.042	0.127	0.047				
3	0.041	0.037	0.035	0.118				

Measure: MEASURE 1

Pairwise Comparisons

(1)	(J)	Mean	,		Differe	nce(a)
regularization	regularization	Difference (I J)	Std. Error	Sig.(a)	Lower Bound	Upper Bound
1	2	,039	,062	1,000	,119	,197
	3	,043	, <mark>05</mark> 1	1,000	,174	,089
2	. 1	,039	,062	1,000	,197	,119
	3	,082	,074	, <mark>846</mark>	,271	,108
3	/ 1	,043	,051	1,000	,089	,174
	2	. ,082	,074	,846	,108	,271

a. Adjustment for multiple comparisons: Bonferroni.

11. filter * regularization

Measure: MEASURE 1								
				95% Confide	ence Interval			
filter	regularization	Mean	Std. Error	Lower Bound	Upper Bound			
1	1	0.577	0.061	0.704	0.451			
	2	0.740	0.083	0.911	0.568			
	3	0.364	0.079	0.526	0.202			
2	1	0.575	0.050	0.471	0.679			
	2	0.660	0.084	0.487	0.832			
	3	0.447	0.083	0.277	0.617			

SPSS output: Pairwise comparisons part II (shortened) - Listening test II.
Instructions

Liebe/r Versuchsteilnehmer/in,

in diesem Versuch wird ein realer Lautsprecher mit einer über Kopfhörer eingespielten Simulation desselben Lautprechers direkt verglichen. Ziel ist es, die Simulation zu erkennen und klangliche Abweichungen der Simulation von der Realität zu bewerten. Dazu wirst du verschiedene Hörbeispiele hören, die sich – auch untereinander – klanglich unterscheiden. Zudem werden die Beispiele mit unterschiedlichen Inhalten (Rauschen und akustische Gitarre) präsentiert.

Die verschiedenen Simulationen werden paarweise im direkten Vergleich zum Schallfeld des Lautsprechers dargeboten und können über die beiden Play-Tasten abgespielt werden. Deine Aufgabe ist es,

- 1.) wenn möglich, die Simulation zu erkennen und dann
- 2.) anhand des zur Simulation gehörenden Schiebereglers zu beurteilen, wie stark sich diese nach deiner Meinung vom Schallfeld des realen Lautsprechers unterscheidet (von "identisch" bis "sehr unterschiedlich").



Bei letzterem ist vor allem darauf zu achten, dass die gegebenen Bewertungen auch die Unterschiede zwischen den einzelnen Simulationen widerspiegeln, so dass sich eine Abstufung der Ähnlichkeit ergibt.

Um die Erkennung zu erleichtern, kannst du die reale Lautsprecherwiedergabe jederzeit mit der **Ref**-Taste aktivieren. Zur Bewertung können alle Klänge beliebig oft angehört werden. Da die Simulation interaktiv auf Kopfbewegungen reagiert, kannst du, z. B. um die Unterschiede besser zu erkennen, deinen Kopf in der Horizontalebene drehen (\pm 80°).

Der Versuch beginnt mit einer kurzen Trainingsphase, um dich mit den auftretenden klanglichen Unterschieden und dem Bewertungsprozess vertraut zu machen. Sollte es noch Unklarheiten geben, kannst du während des Trainings gerne Fragen stellen.

Questionnaire

Abschließend bitte ich dich noch um einige persöniche Angaben. Diese sind für das Experiment genau so wichtig, wie deine Bewertungen.

Name oder Akronym:				
Geschlecht:		w eiblich		m ännlich
Alter:		Jahre		
Bist du musikalisch ausgebildet?		nein		j a
Falls ja, welches Instrument spielst du bzw. welche Art der musikalischen Ausbildung hast du gehabt?			,	Dauer: Jahre
			,	Dauer: Jahre
			,	Dauer: Jahre
Hast du Erfahrung mit Hörversuchen?		n e i n		ja
Bitte nenne die Merkmale anhand derer du Simu konntest. Falls es mehrere Merkmale waren, brir dem offensichtlichsten:	latio nge si	n und Referenz vone e bitte in eine Reihe	inan nfolg	der unterscheiden ge, beginnend mit
Weitere Bemerkungen:				

Appendix C

Other

Ear mould crafting



Custom ear moulds from the archive of the Audio Communication Group.

Appendix D

Electronic documentation

If not appended contact Fabian Brinkmann: brinkmann.f ät gmail.com