Audio Engineering Society

Convention Paper

Presented at the 126th Convention
2009 May 7–10 Munich, Germany

The papers at this Convention have been selected on the basis of a submitted abstract and extended precis that have been peer reviewed by at least two qualified anonymous reviewers. This convention paper has been reproduced from the author's advance manuscript, without editing, corrections, or consideration by the Review Board. The AES takes no responsibility for the contents. Additional papers may be obtained by sending request and remittance to Audio Engineering Society, 60 East 42nd Street, New York, New York 10165-2520, USA; also see www.aes.org. All rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

Evaluation of Equalization Methods for Binaural Signals

Zora Schärer[1] and Alexander Lindau[1]

[1] Audio Communication Group, TU Berlin,
Sekr. EN-08, Einsteinufer 17c, 10587 Berlin, Germany

zora.schaerer@tu-berlin.de, alexander.lindau@tu-berlin.de

ABSTRACT

The most demanding test criterion for the quality of binaural simulations of acoustical environments is whether they can be perceptually distinguished from a real sound field or not. If the simulation provides a natural interaction and sufficient spatial resolution, differences are predominantly perceived in terms of spectral distortions due to a non-perfect equalization of the transfer functions of the recording and reproduction systems (dummy head microphones, headphones). In order to evaluate different compensation methods, several headphone transfer functions were measured on a dummy head. Based upon these measurements, the performance of different inverse filtering techniques re-implemented from literature was evaluated using auditory measures for spectral differences. Additionally, an ABC/HR listening test was conducted, using two different headphones and two different audio stimuli (pink noise, acoustical guitar). In the listening test, a real loudspeaker was directly compared to a binaural simulation with high spatial resolution, which was compensated using seven different equalization methods.

1. INTRODUCTION

Dynamic binaural synthesis is a powerful tool for the simulation of acoustical environments. The movements of a listener's head are detected with a head tracker while an anechoic signal is convolved with a binaural room impulse response (BRIR) matching the head positions. The audio signals are usually reproduced over headphones, but also transaural approaches exist [1].

For an authentic simulation, the auralization system should be transparent, i.e. the synthesized signal should be indistinguishable from the natural sound field. Listening tests on the plausibility of binaural simulations have shown that spectral distortions are the most obvious indication for differences [2][3]. Therefore, one main concern is to optimize the compensation of the electro-acoustic transducers involved in the binaural simulation system. Primarily, the loudspeaker and the dummy head microphones used for BRIR recording and

the headphone used for reproduction need to be considered (Figure 1).

1.1. Headphone Reproduction

When evaluating headphone compensation, it should be kept in mind that the transfer function of the headphone comprises of the transfer function of the transducer itself *and* of the transfer function from the transducer to the individual's ear canal [4]. Appropriate design goals for headphone frequency responses have been proposed by numerous authors [5][6][7][8]. Depending on the application, free- resp. diffuse-field calibrated headphones are common today. However, according to measurements in [9], differences in frequency responses of different headphones can be as large as variations in head related transfer functions (HRTFs). Furthermore, differences can be induced by the selected measurement method, i.e. whether a coupler is used, a dummy head's or a real person's ear [10]. Depending on the type of headphone (extraaural, circumaural, supraaural, suspended on concha etc.), leakage due to small air gaps was shown to contribute to variability in the low frequency response [10]. In [9] it was shown that with blocked ear canal measurements, headphone responses measured on 40 different subjects exhibited common structures up to 12 kHz, whereas with an open ear canal measurement results diverged noticeably above 2 kHz. This finding was related to individual acoustical loading presented by the open ear canal. The variability of measurements of headphone transfer functions for different subjects was examined in more detail in [9][11]. For a single headphone, both studies revealed interindividual differences up to ±10 dB. Therefore, individual headphone compensation was recommended for binaural reproduction by different authors [9][11][12][13].
Apart from interindividual variations of HRTFs and the variations of different headphones' transfer functions, the variability only due to repeated placement of headphones on the same subject was investigated in [4][14]. For a supraaural headphone, differences of ±4dB below and up to ±10dB above 10 kHz were reported in [14]. In [4], largest differences were reported below 500 Hz and above 10 kHz. In [14] it was shown that the result of an equalization based on a single transducer measurement without repositioning can become worse than having no compensation at all. Inverse filters should therefore be derived from an average of multiple measurements carried out while successively repositioning the headphones.

In the quantitative part of this study seven different types of headphones were included. In the listening test two circumaural headphones were evaluated. All measurements were conducted at the blocked ear canal using the FABIAN head and torso simulator (HATS), whose outer ear and head are made from an individual's moulds [3].

1.2. Frequency Response Compensation

BRIRs measured in a more or less diffuse field include the frequency-dependent directional behavior of the measurement loudspeaker, which cannot be equalized with a monaural equalization filter. Instead, the influence of the loudspeaker would be best compensated by adjusting its directivity to the sound source to be simulated [15]. The remaining transducers can be equalized with a digital filter H_c (Figure 1) calculated on the basis of measured transfer functions. If the frequency responses of the dummy head microphones and the headphones are measured simultaneously, they can be equalized with a single filter [16].

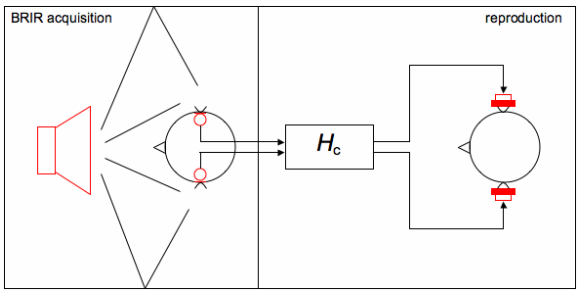

Figure 1: Electro-acoustic transmission in a binaural simulation system: BRIR measurement with dummy head (left) and reproduction with headphones (right), equalized with a compensation filter H_c.

Different techniques for the design of digital inverse filters have been proposed. In the context of loudspeaker equalization some of these methods were evaluated by subjective listening tests [17]. Design techniques in the frequency and time domain were investigated for effects of different filter lengths as well as parameters concerning regularization in the least squares approach (see section 2.3). Additionally, the impact of an off-axis listener position, which in a way is comparable to repositioning effects of headphones, was evaluated and different loudspeakers were compared. It was observed that the performance of an equalization filter can strongly depend on the loudspeaker (or room) to be corrected [17][18]. Instead of conducting listening tests, other evaluations of compensation methods assessed the filter performance descriptively [19] or

calculated quantitative error measures in the frequency and time domain [18]. In some cases, also psychoacoustical criteria were applied to the spectral and time domain behavior of the equalized transfer functions [20].

While the studies mentioned above were focused on the compensation of loudspeakers and room correction, for the equalization of binaural signals, the differences between typical loudspeaker and headphone transfer functions as well as changes in the frequency response due to varying headphone positions and individual transfer paths to the eardrum have to be taken into account. Using quantitative auditory measures and a listening test, we have compared different compensation methods, based on non-individually measured headphone transfer functions, in terms of their performance in the context of dynamic binaural synthesis, with respect to their robustness towards varying headphone positions and to the computational cost.

2. EQUALIZATION

2.1. Headphone Transfer Functions

Figure 3 shows the transfer functions of seven exemplary headphones measured on the FABIAN HATS, usually used by us for BRIR acquisition [3]. Therefore, all headphone transfer functions shown here still contain the frequency response of the dummy head microphones, as these influences have to be compensated, too (Figure 2). Measurements were conducted in the anechoic chamber of the TU Berlin using a 'Monkey Forest' [21] measurement system. It was operated at a sampling rate of 44.1 kHz using a 2^{16} samples sine sweep to provide a good signal-to-noise ratio. All headphones were measured ten times each, while being repositioned on the dummy head between measurements.

The selected headphones can be distinguished with respect to their transducer principle, acoustic coupling and design of the earphones. Table 1 shows an overview of the specifications. For an unobstructed reproduction of the binaurally simulated ear signals, headphones with free-air equivalent coupling (FEC) should be used [16]. With FEC headphones it is ensured that the acoustic radiation impedance as seen from the entrance of the ear canal with an earphone placed on the ear is equivalent to the impedance in the absence of an earphone, i.e. it is equivalent to the impedance of the free sound field.

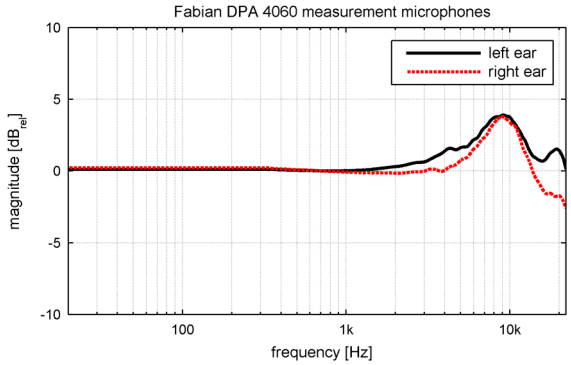

Figure 2: Frequency responses of measurement microphones used with FABIAN (DPA 4060)

Only then, the resulting sound pressure at the entrance of the ear canal and thereby also at the ear drum [16] can be equal for both situations. Since the Stax headphones and the AKG K-1000 meet this requirement almost perfectly [9], they are of particular interest in the context of this study.

Headphone (Preamplifier)	Trancducer Type	Acoustic Coupling	Design
Audio-Technica ATH-M40fs	electro-dynamic	closed	circumaural
AKG K-401	electro-dynamic	half-open	circumaural
Sennheiser Headset H410	electro-dynamic	half-open	supraaural
Stax Lambda Pro New (SRM 313)	electro-static	open	circumaural
Stax SRS 2020 Lambda Basic (SRM 212)	electro-static	open	circumaural
Stax SRS 2050 II (SRM 252 II)	electro-static	open	circumaural
AKG K-1000 (Carver PM)	electro-dynamic	open	extraaural

Table 1: Specifications of the measured headphones

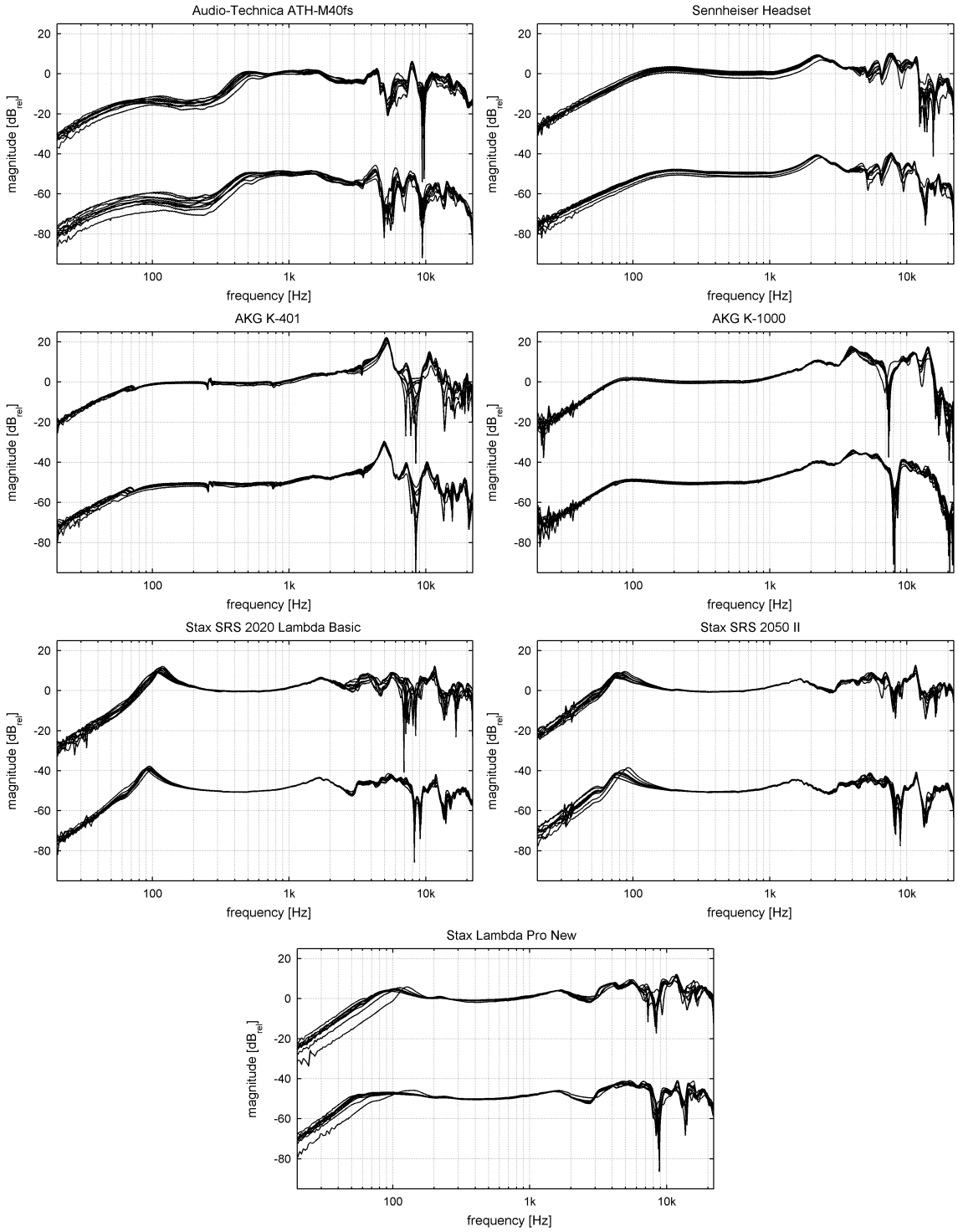

Figure 3: Transfer functions of seven headphones, each measured ten times after repositioning (right channel is shifted by -50 dB for convenience)

The diffuse field equalization of headphones proposed as a standard in [7] minimizes distortion of the sound image when stereo signals intended for loudspeaker reproduction are listened to with headphones. This can be achieved by matching the frequency response of the headphone to an HRTF averaged over several directions of the incident sound. With a consequent implementation of this standardization, a single correction filter could be used for several headphones. However, the large differences between headphones, as seen in Figure 3, show that an individual compensation for each headphone is indispensable in practice.

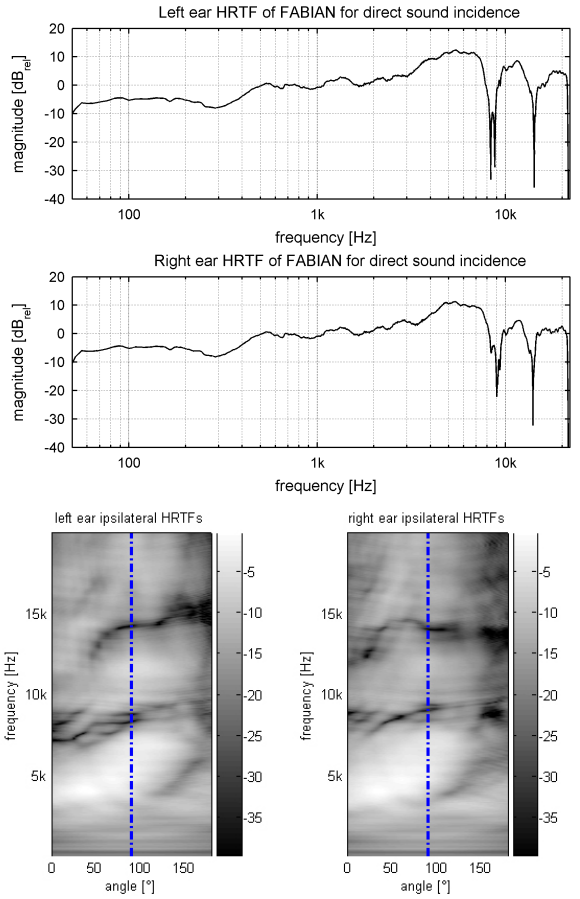

Figure 4: Above: FABIAN's left and right ear's direct incidence HRTF, prominent notches at 8.5 kHz resp. 9 kHz and 14 kHz are clearly visible; below: both HRTFs for the horizontal, ipsilateral hemisphere, an angle of 0° indicates frontal sound incidence, 90° sound incidence (corresponding to picture above) is indicated by broken line

The transfer functions of headphones show considerable changes depending on their position on a listener's head, as illustrated by the ten measurements for each headphone (Figure 3). Even though these measurements were conducted on the same dummy head, the repositioning resulted in shifts of narrow high-frequency notches. For all headphones, except the supraaural headset, these notches are located around 8.5 kHz (left ear) resp. 9 kHz (right ear). As these frequencies are related to a wavelength of ca. 4 cm, they could be due to a destructive interference inside the cavum conchae [22]. This assumption is supported when looking at FABIAN's HRTFs for sound incidence directly from the side, which should be comparable to the headphone reproduction, at least for the circumaural headphones (Figure 4, upper plots).

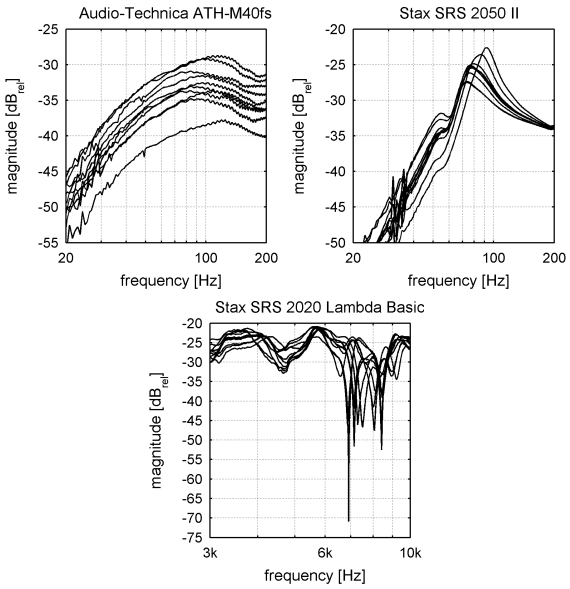

Figure 5: Examples of variations of headphone transfer functions caused by repositioning: low frequency range of the Audio-Technica ATH-M40fs (top left) and the Stax SRS 2050 II (top right) and high frequency range of the Stax SRS 2020 Lambda Basic (bottom)

The second notch in Figure 4 (top) at 14 kHz can also be found in most of the headphone responses (Sennheiser Headset, AKG K401, and all Stax headphones), although it is somewhat attenuated. The lower plots in Figure 4 reveal the cavum conchae notches to be present for sound incidence from most of the ipsilateral hemisphere (horizontal plane). The lower notch around 8 kHz is present for ipsilateral sound incidence from 0-100°, the second notch at about 14 kHz occurs for sound

incidence from 50° to 180°. The AKG K-1000, an extraaural headphone, was measured with its transducers adjusted to an opening angle corresponding to 55° sound incidence; as can be seen this eliminates the second notch completely in the right ear's headphone response (Figure 3).

In most cases, the low-frequency roll-off of the headphone transfer functions was affected by the repositioning, too. However, the AKG K-1000 exhibited a robust frequency response in a wide range of the pass-band. This may be an indication for the positive effect of extraaural earphones, which was also observed in [10]. Figure 5 shows some detailed examples of the typical variations caused by changing headphone positions for the Audio-Technica ATH-M40fs, the Stax SRS 2050 II and the Stax SRS 2020 Lambda Basic.

2.2. Inverse Filtering

In order to linearize a transfer function $H(k)$, a filter $H_c(k)$ needs to be calculated that fulfills:

$$H_{eq}(k) = H(k) \cdot H_c(k) = 1 \quad (1)$$

In the time domain this corresponds to a convolution with the filter's impulse response h_c, resulting in a system response that is a dirac pulse:

$$h_{eq}(n) = h(n) * h_c(n) = \delta(n) \quad (2)$$

The direct inversion of the measured frequency response, i.e.

$$H_c(k) = \frac{1}{H(k)} \quad (3)$$

comes up against certain constraints when dealing with electro-acoustic (mixed-phase) systems:

- The direct inversion of mixed-phase systems yields an acausal, infinite and potentially instable impulse response [23].

This can be avoided by splitting the transfer function into a minimum-phase and an allpass component, of which only the former is equalized [24]. This leads to an exact compensation of the amplitude but not of the phase response. However, in most cases the result of this method is insufficient due to the remaining error energy in the allpass component [17][24]. An alternative approach is the homomorphic decomposition of $h(n)$. By means of cepstral analysis, the minimum- and maximum-phase components of the system response are derived and inverted separately [18]. It has been shown in [18] though, that the homomorphic decomposition is prone to numerical errors and therefore not suitable either. Instead, the most simple and effective way to overcome the described problem is to delay the target function of the compensation, usually by half the length of the filter impulse response. By introducing such a modeling delay, the acausal part of the filter impulse response is shifted into the positive part of the time domain [24][25][26]. This approach was adopted here.

- A compensation of the complete frequency range would lead to excessive boosting beyond the high and low roll-off frequencies (see Figure 3).

In order to avoid this problem it is necessary to specify an appropriate target function for compensation. The delta function in eq. (2) is thereby replaced by the impulse response of a bandpass filter, limiting the out-of-band amplification in the compensation filter. As this bandpass filter might cause perceivable group delay distortions, a linear-phase FIR-filter was chosen. After inspecting the measured headphone responses, the target bandpass was defined to have a pass-band between 50 Hz and 21 kHz; its stop-band attenuation was set to 60 dB. With a linear-phase bandpass, the modeling delay described above is already part of the target function, since every linear-phase filter is characterized by a half-length delay.

- Perfect compensation is restricted to a specific and time-invariant position of source (headphone) and receiver (dummy head microphone or listener's ear), which can be described by an individual transfer function for every configuration.

This problem has already been illustrated by the measurements shown above. As also reported in [4], the variation of headphone positions is particularly critical for the compensation of the upper frequency range. An inverse filter calculated from eq. (3) compensates narrow high-frequency notches with corresponding peaks. Any perfect compensation would then after repositioning lead to ringing artifacts due to now slightly 'mistuned' peaks of the inverse filter. A common approach to overcome this problem involves a reduced compensation accuracy of the deep cavum conchae notches. Although these are still present in the equalized

transfer function, the result of the compensation is less affected by artifacts caused by changing source/receiver positions. This approach is based on the fact that deep notches are less disturbing than narrow peaks from a perceptual point of view [27]. Additionally, the inverse filter should be calculated from an average of several transfer functions [14], also resolving the restriction of the compensation to a specific frequency response. Moreover, this approach causes narrow notches to be averaged out prior to the actual computation of the filter. Hence, all filter design methods described below were based on a complex average of the ten measurements conducted for each headphone.

2.3. Methods

Several filter design techniques for equalization were chosen from literature and implemented in Matlab®. Depending on the strategy to reduce accuracy in the various methods, they can be divided into two groups (see Figure 6).

The first approach uses a *preprocessed transfer function* for inversion, taking into account the requirements related to target function and modeling delay. The preprocessing aims at simplifying the measured frequency response, which will in turn lead to a non-perfect equalization filter after inversion. This can be achieved by magnitude or complex smoothing as described in [28].

$$\left|H_{\mathrm{sm}}(k)\right| = \sqrt{\sum_{i=0}^{N-1}\left|H\left[(k-i) \bmod N\right]\right|^2 \cdot W(m,i)} \quad (4)$$

$$H_{\mathrm{sc}}(k) = \sum_{i=0}^{N-1} H\left[(k-i) \bmod N\right] \cdot W(m,i) \quad (5)$$

The smoothed transfer function is derived by circular convolution of the measured transfer function with a smoothing window $W(m,k)$. m denotes the width of this window, which grows with increasing frequency index k. In the case of complex smoothing, this is equivalent to apply time domain windowing with window lengths inversely proportional to frequency. Despite being computationally more expensive when compared to classical logarithmic smoothing of the magnitude spectrum, this method has the advantage to preserve low frequency resolution while still effectively windowing the high frequency transients [28]. The resolution of the smoothing function can be defined freely. For example, a fractional octave band or an ERB scale may be applied. Here, octave band smoothing was employed.

The *compare and squeeze* technique [21] uses a slightly different method for preprocessing. Here, the smoothed frequency response is fitted to the original response with a predefined weighting. Peaks are usually matched perfectly whereas notches remain smoothed to a certain extent, the latter therefore being compensated with reduced accuracy after inversion.

Finally, the measured transfer function can also be approximated by an *all-pole model*, where resonances are represented accurately as opposed to notches [29]:

$$H_{\mathrm{p}}(z) = \frac{G}{1 + \sum_{l=1}^{P_i} d_l z^{-l}} \quad (6)$$

G is a constant scaling factor, the coefficients d_l are calculated with the least squares method [30]. The advantage of an all-pole model is that its inverse can be readily derived by inverting eq. (6). Furthermore, as the resulting expression is a non-rational polynomial in z, the system will always be stable and causal.

The second group of filter design techniques is based on a *least squares criterion*, minimizing the error between the target function and the equalized frequency response. By employing frequency-dependent regularization, the effort towards error minimization can be controlled, i.e. the accuracy of the compensation is manipulated in specific regions of the transfer function. Using matrix notation, the computation of the equalization filter in the time domain is given by the following expression [26]:

$$\mathbf{h}_{\mathrm{c}} = \left[\mathbf{H}^T\mathbf{H} + \beta \cdot \mathbf{B}^T\mathbf{B}\right]^{-1} \cdot \mathbf{H}^T\mathbf{d} \quad (7)$$

H and **B** are convolution matrices of the measured transfer function and the regularization filter respectively. $\mathbf{h}_{\mathrm{c}}$ and **d** are signal vectors, the latter denoting the (delayed) impulse response of the target bandpass filter. β is a scalar additionally weighting the regularization filter. The same calculation can be done in the frequency domain, the corresponding expression from [17][31] is:

$$H_c(k) = \frac{D^*(k)H(k)}{H(k)H^*(k) + \beta \cdot B(k)B^*(k)} \quad (8)$$

It has been shown in [32] that regularization helps to reduce temporal aliasing artifacts in the equalized impulse response. Temporal aliasing or wrapping artifacts are caused by a circular convolution involved when calculating the compensation filter in the frequency domain and can also be reduced by windowing the inverse filter [32], as has been done here. When correction filters are calculated using regularization, the result of the compensation will be least accurate in the pass-band of the regularization filter employed. Hence, a highpass filter [33] was used for the equalization of the headphones in this study. Thus, the effort for compensation in the high frequency range, where deep cavum conchae notches are most prominent, was restricted. The filter cut-off frequency was at about 8 kHz with a wide transition range to -20 dB in the stop-band. Additionally, an octave band smoothed inverse of the headphone transfer function was used as an alternative regularization filter, similar to [17][20] and [34], which allowed for an individual adjustment of the equalization accuracy depending on the existing notches. For both filters, the weighting term β was determined by three expert listeners in a preliminary listening test conducted in accordance with the main test described in section 3.2. The values were set to $\beta = 0.4$ for the highpass and $\beta = 0.07$ for the smoothed, inversed filter. It should be emphasized that suitable values for β are heavily dependent on the frequency response to be inverted.

Figure 6 gives an overview of the seven compensation methods examined in this study. They are ordered according to the specific approaches used. All techniques described above yield FIR equalization filters. Therefore, the filter length is another design parameter, which influences the filter performance. The length of all seven filters was set to 2048 samples. This parameter was not varied because shorter lengths would yield an insufficient filter performance [17]. Higher FIR filter orders were not suitable either, as the filters were linear-phase and the resulting predelays lead to audible latency in the binaural simulation. The modeling delay in the filter impulse response directly increases the duration of the cross-fade of BRIRs in the dynamic auralization. Hence, the maximum linear-phase filter length for compensation should be adapted to a threshold of just noticeable latency in binaural simulations. A recent study on this subject has been presented in [35]. For a discussion on minimum- vs. linear-phase target function, see section 6.

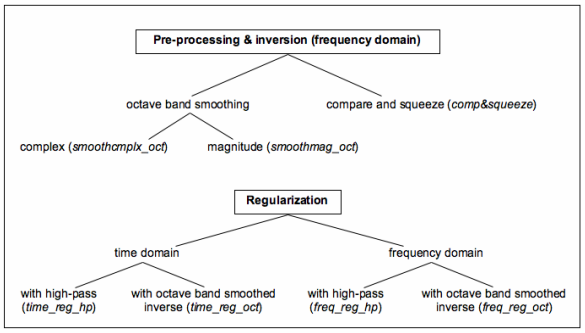

Figure 6: The seven inverse filter design techniques examined in this study; the terms in brackets indicate the abbreviations used below.

3. EVALUATION

3.1. Validation and Auditory Simulation

In order to assess the performance of the examined filter methods and to determine certain filter parameters (e.g. bandwidth of smoothing) prior to the listening test, a frequency-dependent error measure was calculated considering the pass-band region of the target bandpass filter. At first, the equalization result was simulated by taking into account the varying headphone positions. The impulse responses of the compensation filters were thereby convolved with each of the ten measured headphone impulse responses. The power spectra of the target bandpass and the equalized transfer functions were then analyzed with an auditory filter bank of 40 overlapping equivalent rectangular bandwidth (ERB) bandpass filters emulating the frequency selectivity of the auditory mechanism [36]. The error was calculated in each band as the difference between the two spectra [37]:

$$E(f_c) = 10\log\left(\frac{\int_{-\infty}^{\infty} C(f, f_c)\left|H_{eq}(f)\right|^2 df}{\int_{-\infty}^{\infty} C(f, f_c)\left|D(f)\right|^2 df}\right) \quad (9)$$

$C(f,f_c)$ is an auditory filter with the center frequency f_c, $H_{eq}(f)$ and $D(f)$ denote the equalized and the bandpass

frequency response respectively. A perceivable difference was assumed for a deviation greater than 1 dB.

According to this preliminary analysis, the all-pole method proved to be unsuitable for the equalization of the headphones. This was related to the problem of determining an appropriate order of the prediction model. For small orders, the low frequency resonance of the Stax SRS 2050 II was not represented adequately and hence compensated insufficiently. On the other hand, high-frequency notches were modeled with great accuracy when the order was increased. This caused undesired peaks in the equalized transfer function, particularly of the Stax Lambda Pro New. As a suitable frequency-dependent parameterization of the all-pole approach seemed not readily available, it was discarded from further consideration.

Figure 7 and Figure 9 show two examples for the expected compensation results, simulated by convolving a correction filter with the ten original measurements of the Stax SRS 2050 II and the Stax Lambda Pro New respectively. In both cases, a filter with highpass regularization calculated in the frequency domain (*freq_reg_hp*) with a length of 2048 samples was used. This equalization method yielded particularly good results in the listening test (see section 4). The frequency-dependent error measure, $E(f_c)$, computed for each of the ten simulated equalization results is depicted in Figure 6 for the Stax SRS 2050 II and in Figure 8 for the Stax Lambda Pro New. Again, the results reveal the problem of variations in the frequency responses caused by repositioning of the headphones. The low-frequency roll-off region (50-160 Hz) is highly variable and shows shifted peaks, which were clearly audible according to results from the ERB analyses and the listening test. The compensation result of the Stax SRS II 2050 II is quite robust towards variations of positioning in the pass-band. However, above 3 kHz the deviations increase for both headphones. Particularly the equalized frequency response of the Stax Lambda Pro New exhibits a highly unsystematic distribution of the error measure along with audible differences in the pass-band. A validation of these simulations is shown in Figure 11. Five transfer function measurements with repositioning in between were conducted on the dummy head for both headphones. The measurement stimulus was pre-filtered with the impulse response of the compensation filter (*freq_reg_hp*), the measured transfer functions thus reveal the actual performance of the equalization.

The validation measurements evidently confirm the simulations shown above. Particularly the dramatic variations in the low frequency range are verified.

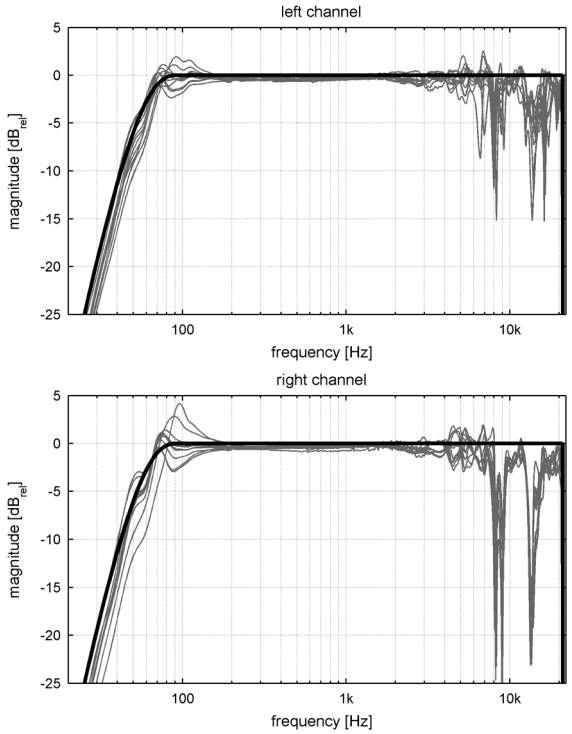

Figure 7: Ten measurements of the Stax SRS 2050 II equalized with *freq_reg_hp*, left channel (top) and right channel (bottom), heavy black line: target bandpass

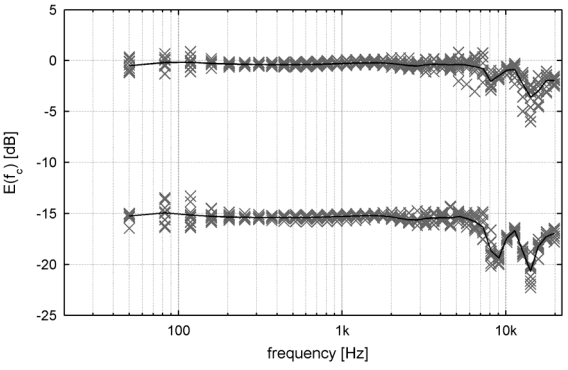

Figure 8: Frequency-dependent error of the Stax SRS 2050 II compensation, derived from the ERB filer bank analysis for left channel (top) and right channel (bottom, shifted -15 dB), black line: mean error

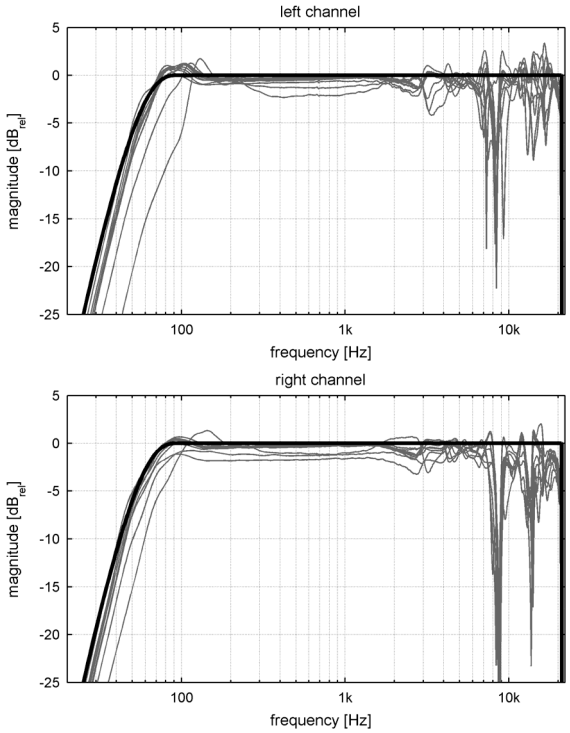

Figure 9: Ten measurements of the Stax Lambda Pro New equalized with *freq_reg_hp*, left channel (top) and right channel (bottom), black line: target bandpass

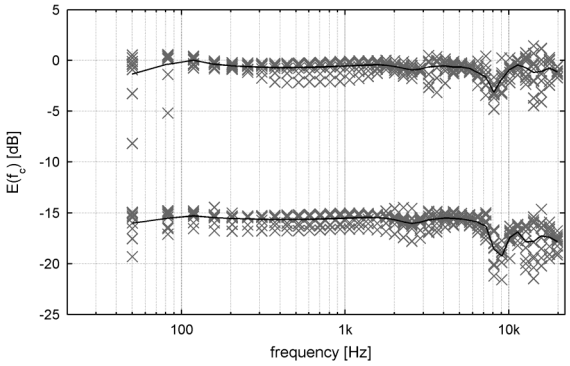

Figure 10: Frequency-dependent error of the Stax Lambda Pro New compensation, derived from the ERB filer bank analysis for left channel (top) and right channel (bottom, shifted -15 dB), black line: mean error

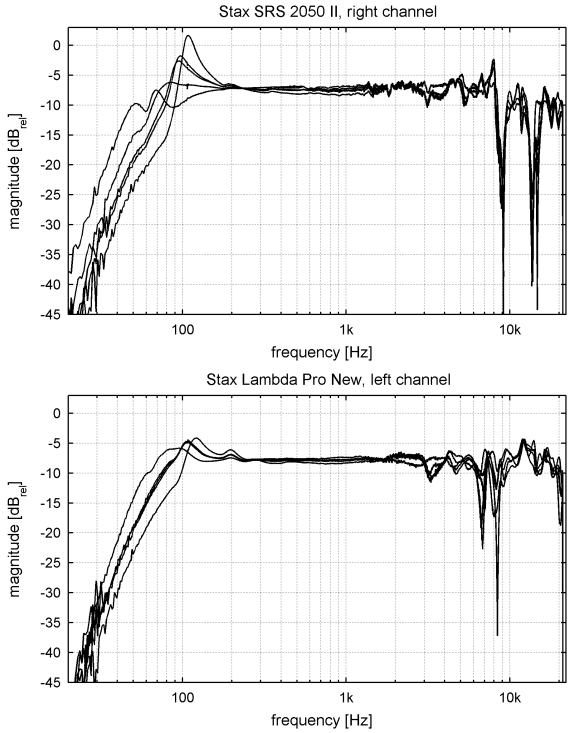

Figure 11: Validating compensation results by measurement. Five measurements of the compensation with *freq_reg_hp*; top: Stax SRS 2050 II (right channel) and bottom: Stax Lambda Pro New (left channel).

3.2. Listening Test

In order to determine a perceptually superior filter design technique for the equalization of binaural signals, the similarity of a simulated and a real sound source was rated using direct comparison in a listening test.

3.2.1. Auralization

The listening test took place in the Electronic Studio of the TU Berlin ($V = 160$ m^3, $RT = 0.7$ s, Figure 12). The simulated sound source was a loudspeaker (Meyer Sound UPL-1) placed in front of the listener at a distance of 2.9 m and a height of 2 m. The binaural simulation was therefore based on a set of BRIRs measured for a discrete grid of head positions with the FABIAN HATS while using the Meyer Sound loudspeaker. The BRIRs were measured with an inaudibly fine angular resolution of 1° for a ±65° horizontal and -40°/+30° vertical range of head movements [38]. Thus, the dataset to be auralized consisted of 9301 BRIRs. The listening test was conducted using the two head-

phones discussed above (Stax SRS 2050 II and Stax Lambda Pro New), as they were considered to be FEC compatible. Additionally, these headphones exhibit only minor obscuration of the outer sound field, so subjects kept wearing the acoustically transparent headphones throughout the listening test, while listening either to the simulation or to the real loudspeaker in front of them. To maintain identical reflection and diffraction characteristics around the head for the real and the simulated sound field during the listening test, the headphones were placed on the dummy head during the measurement. As both headphones are of identical outer dimensions, only one set of BRIRs had to be measured.

Figure 12: Listening test environment with the reference loudspeaker from the perspective of the listener

The auralization system described in [33] was used for the dynamic binaural synthesis of the studio environment.

3.2.2. Listening Test Design

The listening test was an ABC/HR design (slightly modified from [39]). Here, a stimulus marked as reference (A) – the real loudspeaker – is compared to two other stimuli (B and C). One of these signals corresponded to the reference again, while the other one was a binaural simulation of the loudspeaker presented over headphones. The latter were equalized with one of the seven compensation methods described above (see Figure 6). Subjects were requested to rate the perceived similarity between the equalized simulation and the natural sound field on a scale from "identical" to "very different" (Figure 13). Thus, before starting to rate, subjects had to decide which of the alternatives B and C was identical to the reference. So, inherently it was tested if they were actually able to detect the real sound source. All seven equalization methods could be selected one at a time. They were interchanged without audible artifacts by means of an additional convolution plug-in employed prior to the actual auralization engine. Additionally, an uncompensated version of the simulation was added to the test as a hidden anchor.

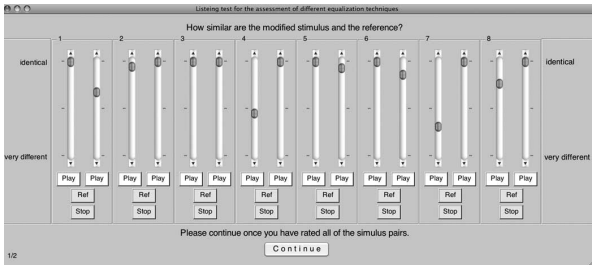

Figure 13: Graphical user interface of the listening test

The evaluation of all equalization filters was conducted for the two headphones (Stax SRS 2050 II and Stax Lambda Pro New) and two audio samples (pink noise and acoustic guitar). All subjects had to rate all possible combinations of stimuli resulting in a full factorial 7x2x2 test design with repeated measurements. The stimulus duration was 5 seconds and 20 ms raised cosine fade-ins and -outs had been applied. The audio samples were filtered with the target bandpass used for the headphone compensation in order to remove excessive low-frequency energy, as this would have been an obvious cue for discriminating between the simulated and the real sound source. The listeners were presented with the graphical user interface shown in Figure 13 and completed a total number of four runs. In every run, the seven randomly ordered filters and the hidden anchor could be selected and compared as often as desired. Thus, the ratings of the compensation methods were assured to be relative to each other while regarding a single audio content and one headphone each. For each subject the sequence of combinations of audio samples and headphones to be rated was randomized.

Apart from exchanging the headphones after a rating run they were not taken off during the whole procedure (see section 3.2.1). The seating position of the listeners was kept in accordance with the position of the HATS during the BRIR measurement using visual guidance. The correction filters were calibrated to equal loudness using an algorithm described in [40]. The loudness adjustment of the headphone and the loudspeaker signal was then accomplished by two expert listeners.

The listeners completed a training session in the beginning, to get used to the rating process and the range of typical compensation artifacts. After the listening test, subjects were requested to fill in a questionnaire indicating qualities that were perceived relevant for the discrimination of the real and simulated sound source.

3.2.3. Subjects and Sampling

28 subjects aged from 24 to 43 took part in the listening test, 25 were male and 3 female. Most of them had experience with listening tests and could be considered as trained listeners.

Since the suitability of a particular equalization method for a particular headphone would correspond to a significant first order interaction in the three factorial test design, and since this effect size was assumed to be small, the optimal sample size was calculated to be $N = 26$, following [41]. In this way a minimum sample defined which guaranteed to reliably detect a meaningful effect.

4. RESULTS

Figure 14 shows an overview of the attributes given by the listeners as indicators for differences of the simulated to the real sound field. Attributes regarding coloration or timbre were given most frequently and even exclusively by many subjects. Apparently, perceived boosting of high frequencies as well as ringing artifacts were more dominant than the missing bass. The perceived difference in the transients of the simulated signal are probably related to the pre-ringing of the linear-phase target bandpass impulse response as well as to the typical artifacts of equalization techniques involving regularization [32]. The latency identified by three listeners goes back to the modeling delay in the correction filters, which directly increases the latency of the dynamic auralization, as stated above. The adequate loudness calibration is confirmed by the fact that loudness differences were only named twice. The rest of the attributes, i.e. spatiality, localization, distance and naturalness, are not necessarily linked to the equalization alone. Diverging localization is also known to arise from ITD mismatch. Errors in perceived distance can be related to non-individual pinnae cues, as distance perception for frontal sound incidence is mainly a matter of learned coloration differences. Spatiality and naturalness seem to be linked to the plausibility of the simulation itself; as these attributes are suspected to be of multidimensional nature, further interpretation is refrained from.

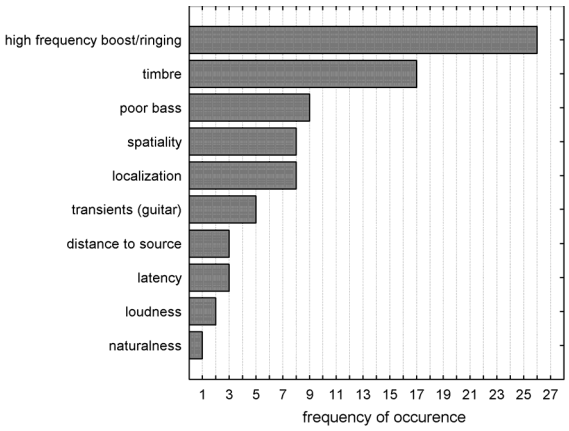

Figure 14: Overview of the attributes of differences between real and simulated sound field presentation given by the listeners

Because of obvious misunderstandings, which were confirmed by an outlier analysis (Grubb's outlier test), the answers of one subject were eliminated from the analysis. Hence, the following results are based on the answers of 27 listeners. A reliability test indicated a high inter-subject consistency with Cronbach's Alpha = 0.955. A Kolmogoroff-Smirnoff test confirmed the normal distribution of the answers for each condition rated in the listening test.

Mean difference grades for the assessment of the equalized binaural stimuli and the reference sound source are shown in Figure 15 with 95% confidence intervals. The *x*-axis is labeled with the abbreviations for the filter design techniques according to Figure 6, *no_filt* marks the hidden anchor (unequalized simulation). Negative difference grades indicate a stimulus performance worse than the reference; positive values would indicate a better performance or more precisely, a situation where the reference situation could not reliably be discriminated from the simulated version.

For a more intuitive interpretation of the results, in Figure 16 difference grades were normalized relative to the rating of the hidden anchor, and converted to percent. This yields a corrected scale ranging from 'absolutely identical' to 'not similar at all', illustrating the filter performance more clearly.

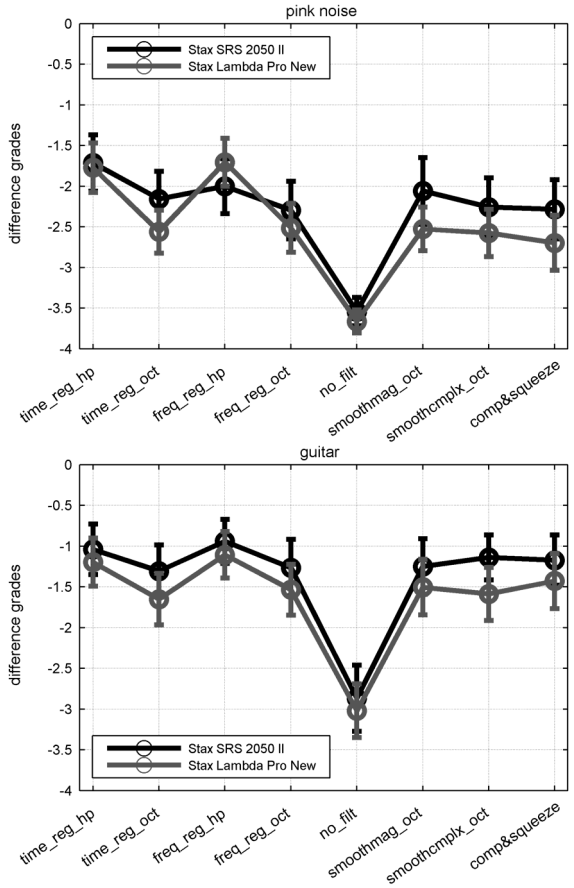

Figure 15: Mean difference grades of the ratings of simulation and reference with 95% confidence intervals for pink noise (top) and acoustic guitar (bottom)

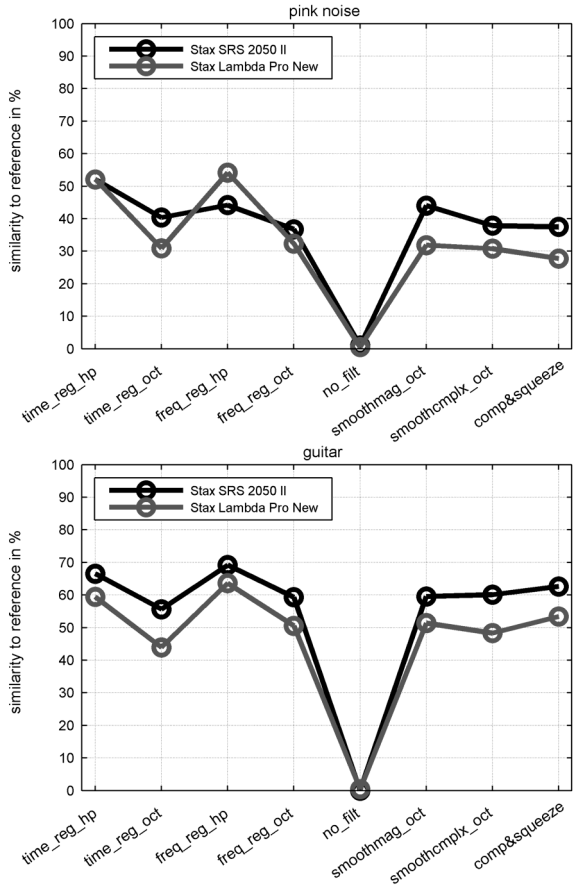

Figure 16: Average perceived concordance of the equalized simulations with a real sound field (normalized to the assessment of the unequalized hidden anchor) for pink noise (top) and acoustic guitar (bottom)

A full factorial 7x2x2 analysis of variance (ANOVA) for repeated measurements was performed on the difference grades obtained from the test data. The preconditions necessary for an ANOVA, such as homogenous error variances and homogenous correlations between conditions on all factor levels, were verified with Mauchly's Test of Sphericity [41]. If necessary, the degrees of freedom for the F-tests were adjusted accordingly.

As the figures immediately reveal, the discrimination between simulation and reality was possible without difficulty for all conditions. However, the simulations processed with a correction filter were always perceived to be more similar to the natural sound field than those with no equalization (*no_filt*). The presentation of the guitar sample yielded significantly higher similarity ratings than pink noise.

This is not surprising, since the spectrum of the natural instrument contains less energy in the critical ranges of the frequency response with respect to varying transfer functions, i.e. at the low and very high frequency end.

With respect to the compensation techniques, the two correction filters designed with highpass regularization evidently received higher ratings than the rest. Pooled over both headphones and audio contents, the ANOVA confirmed the significance of this result but yielded no significant difference between filter design in the frequency or the time domain, even if the latter lead to a slightly superior result. The ratings of all the other correction filters did not differ significantly. However, in the particular case of the Stax SRS 2050 II being equalized with the magnitude smoothing method, an intermediate performance was achieved. Furthermore, it

must be noted that the interaction between the assessment of the compensation methods and the audio content was highly significant. This is illustrated by a comparison of the ratings of the inferior methods with respect to the two best techniques for both audio samples. In the case of the guitar sample, the inferior assessments improved above average.

Except for one condition (*freq_reg_hp*/pink noise), the ratings of the equalized Stax SRS 2050 II were superior to those of the Stax Lambda Pro New with high significance. This might be due to the more pronounced shifts of high frequency notches in the case of the latter headphone (see Figure 9). This assumption is supported by the fact that listeners most frequently named ringing artifacts as attributes for discrimination (Figure 14). However, pooled over both audio contents, the two headphones were assessed equally when being compensated with the methods involving highpass regularization. Thus, even systems with highly varying high-frequency notches, such as the Stax Lambda Pro New, can be corrected adequately as long as the accuracy of the compensation is generally restricted in the high frequency range. The exceptional case mentioned above, where the equalization of the Stax SRS 2050 II with the highpass regularization filter designed in the frequency domain yielded poorer results, might be related to the narrow resonance in the low-frequency roll-off of this headphone (see Figure 11). When employing shorter filter lengths, the equalization of low frequencies becomes more difficult with filters designed in the frequency domain because of the linear resolution of the Fourier transform (see also [17]).

5. DISCUSSION

Several filter design techniques for the computation of inverse headphone filters were evaluated with regard to the compensation of binaural signals. Headphone transfer functions were measured at the blocked ear canal of a dummy head multiple times while being successively repositioned. In a listening test - directly comparing a real and a simulated sound field - two different headphones were examined and their equalization was assessed using two different audio samples. The least squares equalization techniques with highpass regularization were found to be perceptively most efficient. The effect of the stimulus type was obvious, with pink noise being far more critical than a natural musical sound. Regarding the different headphones, results were somewhat ambiguous, though one of the headphones yielded slightly better results, which could be related to a reduced high-frequency variability when being repositioned on the head. The difference between reality and simulation while being given a dedicated reference was easily audible for every combination of stimulus, headphone, and compensation filter. It was argued that this can be mainly assigned to the varying transfer functions caused by repositioning of the headphones. Regarding results from literature, potential errors from using non-individualized BRIRs and non-individualized headphone compensations could have been at least comparable to the variability caused by repositioning (see section 1.1). Consequently, it must be noted that Figure 11 illustrates the best possible equalization, since the corrected transfer function was measured on the ears of the dummy head itself. Thus, our study indicates the maximal performance that can be expected from frequency response compensation when using non-individualized BRIRs *and* non-individualized headphone correction filters. Nonetheless, it was clearly shown that in all cases even a non-individual equalization will yield more plausible simulation results than using no equalization at all.

In the listening test, errors of the compensation were apparent in the high frequency range as well as for the low-frequency roll-off (see Figure 11). In many cases, the missing bass was an obvious cue for the listeners to distinguish between the natural and the simulated sound source. Structure-borne sound transmission, which is principally absent with headphone reproduction, may have supported this impression. As it was not completely controllable, the relative position of the listeners in the sound field may have caused further incongruence between simulation and reality.

6. OUTLOOK

The study revealed several aspects that should be taken into account for future optimization of binaural signal reproduction. Whereas large scale acquisition of individualized BRIR data sets will remain impracticable, individual headphone calibration should instantly yield improvements (see also [13]). Secondly, the design of a headphone more appropriate for compensation in terms of its transfer characteristics should be considered. Such a device would require a largely flat frequency response as well as a position-independent transfer characteristic in the low frequency range. Especially headphones with extraaural earphones have shown to exhibit this characteristic. Additionally, this type of headphone would comply most easily with the FEC criterion. If suspected an issue with conventional headphone transducers, line-

arity over a wide range of the frequency response could be achieved by using a miniature 2-way coaxial system. The reduced low-frequency sensitivity with transaural designs might be compensated for with an additional calibrated subwoofer, which could also deliver the missing impact sound. Further optimization of the regularization method itself should aim at a more detailed examination of the weighting term β and its proper determination for different headphones and each transducer individually. Moreover, a recent study [42] suggested including the regularization directly into a minimum-phase target function. According to informal listening tests [42], this approach reduces the pre-ringing in the corrected impulse response typical for regularization methods using linear-phase target functions, while preserving the perceptual superiority of the approach as approved in our study. In addition, this solution would avoid increasing the overall system latency as - due to the immediate onset of the minimum-phase impulse response - the modeling delay can nearly completely be extinguished after equalization. An evaluation of the audibility of phase errors remaining when using this method has to be conducted still.

7. ACKNOWLEDGEMENTS

This work was supported by a grant from the Deutsche Telekom Laboratories and the Deutsche Forschungsgemeinschaft (DFG).

8. REFERENCES

[1] W. G. Gardner *3-D Audio Using Loudspeakers*. PhD thesis, MIT Media Laboratories, Cambridge (1997)

[2] C. Moldrzyk, T. Lentz, S. Weinzierl „Perzeptive Evaluation binauraler Auralisationen“ DAGA München, In: *Fortschritte der Akustik*, pp. 545-546 (2005)

[3] A. Lindau and S. Weinzierl „FABIAN. Schnelle Erfassung binauraler Raumimpulsantworten in mehreren Freiheitsgraden“ DAGA Stuttgart, In: *Fortschritte der Akustik*, pp. 633-534 (2007)

[4] K. I. Mcanally and R. L. Martin „Variability in the headphone-to-ear-canal transfer function” In: *J. Audio Eng. Soc.*, vol. 50/ 4, pp. 263-266 (2002)

[5] E. A. G. Shaw „Earcanal pressure generated by circumaural and supraaural earphones“ In: *J. Acoust. Soc. Am*., vol. 39, pp. 471-479 (1966)

[6] J. R. Sank „Improved real-ear tests for stereophones“ In: *J. Audio Eng. Soc*., vol. 28/4, pp. 206-218 (1980)

[7] G. Theile „On the standardization of the frequency response of high-quality studio headphones“ In: *J. Audio Eng. Soc*., vol. 34/12, pp. 956-969 (1986)

[8] H. Møller et al „Design criteria for headphones“ In: *J. Audio Eng. Soc*., vol. 43,/4, pp. 218-232 (1995)

[9] H. Møller et al „Transfer characteristics of headphones measured on human ears“ In: *J. Audio Eng. Soc*., vol. 43/4, pp. 203-217 (1995)

[10] F. E. Toole „The acoustics and psychoacoustics of headphones“ In: *Proc. of the 2nd International AES Conference*: *The Art and Technology of Recording*. Anaheim, CA. (1984)

[11] D. Pralong and S. Carlile „The role of individualized headphone calibration for the generation of high fidelity virtual auditory space“ In: *J. Acoust. Soc. Am.*, vol. 100/6, pp. 3785-3793 (1996)

[12] W. L. Martens „Individualized and generalized earphone correction filters for spatial sound reproduction” In: *Proc. of ICAD 2003 - 9th Meeting of the International Conference on Auditory Display*. Boston (2003)

[13] S. Kim and W. Choi: „On the externalization of virtual sound images in headphone reproduction: A Wiener filter approach” In: *J. Acoust. Soc. Am*., vol. 117, No. 6, pp. 3657-3665 (2005)

[14] A. Kulkarni and H. S. Colburn „Variability in the characterization of the headphone transfer-function” In: *J. Acoust. Soc. Am*., vol. 107/2, pp. 1071-1074 (2000)

[15] G. Behler „Variable Richtcharakteristik mit Dodekaeder-Lautsprechern” DAGA Dresden, In: *Fortschritte der Akustik*, pp. 67-68 (2008)

[16] H. Møller „Fundamentals of binaural technology“ In: *Applied Acoustics*, vol. 36, pp. 171-218 (1992)

[17] S. G. Norcross, A. G. Soulodre, M. C. Lavoie „Subjective investigations of inverse filtering“ In: *J. Audio Eng. Soc*., vol. 52/10, pp. 1003-1028 (2004)

[18] J. N. Mourjopoulos, P. M. Clarkson, J. K. Hammond „A comparative study of least-squares and homomorphic techniques for the inversion of mixed phase signals” In: *Proc. of the 1982 IEEE Int. Conference on ASSP*, Paris (1982)

[19] M. Karjalainen, T. Paatero, J. N. Mourjopoulos, P. D. Hatziantoniou „About room response equalization and dereverberation” IEEE Workshop on ASPAA, New Paltz/NY (October 2005)

[20] L. Fielder „Analysis of traditional and reverberation-reducing methods of room equalization” In: *J. Audio Eng. Soc*., vol. 51/2, pp. 3-25 (2003)

[21] S. Müller *Digitale Signalverarbeitung für Lautsprecher* doct. diss., RWTH Aachen (1999)

[22] E. A. Lopez-Poveda and R. Meddis: „A physical model of sound diffraction and reflections in the human concha” In: *J. Acoust. Soc. Am*., vol. 100, No. 5, pp. 3248-3259 (1996)

[23] A. V. Oppenheim, R W. Schafer, J. R. Buck *Discrete-Time Signal Processing*, 2nd Edition, Prentice Hall, Upper Saddle River/NJ (1999)

[24] J. N. Mourjopoulos „Digital equalization of room acoustics“ In: *J. Audio Eng. Soc*., vol. 42/11, pp. 884-900 (1994)

[25] J. N. Mourjopoulos, P. M. Clarkson, J. K. Hammond „Spectral, phase and transient equalization for audio systems” In: *J. Audio Eng. Soc.*, vol. 33/3, pp. 127-132 (1985)

[26] O. Kirkeby and P. A. Nelson „Digital filter design for inversion problems in sound reproduction“ In: *J. Audio Eng. Soc.*, vol. 47/7, pp. 583-595 (1999)

[27] R. Bücklein „The audibility of frequency response irregularities“ In: *J. Audio Eng. Soc*., vol. 29/3, pp. 126–131 (1981)

[28] P. D. Hatziantoniou, J. N. Mourjopoulos, „Generalized fractional-octave smoothing of audio and acoustic responses“ In: *J. Audio Eng. Soc.*, vol. 48/4, pp. 259-280 (2000)

[29] J. N. Mourjopoulos, M. A. Paraskevas „Pole and zero modeling of room transfer functions“ In: *Journal of Sound and Vibration*, vol. 146/2, pp. 281-302 (1991)

[30] M. H. Hayes *Statistical Digital Signal Processing and Modeling*, Wiley, Hoboken/NJ (1996)

[31] O. Kirkeby, P. A. Nelson, H. Hamada, F. Orduna-Bustamante „Fast deconvolution of multichannel systems using regularization“ In: *IEEE Trans. on SAP*, vol. 6/2, pp. 189-194 (March 1998)

[32] S. G. Norcross, A. G. Soulodre, M. C. Lavoie „Distortion audibility in inverse filtering“, *Proc. of the 117th AES Convention,* San Francisco, preprint 6311 (October 2004)

[33] A. Lindau, T. Hohn, S. Weinzierl „Binaural resyhnthesis for comparative studies of acoustical environments“ *Proc. of the 122nd AES Convention*, Vienna, preprint 7032 (May 2007)

[34] S. G. Norcross, M. Bochard, A. G. Soulodre, „Adaptive strategies for inverse filtering“ *Proc. of the 119th AES Convention*, New York, preprint 6563 (October 2005)

[35] A. Lindau, S. Weinzierl „The Perception of System Latency in Dynamic Binaural Synthesis”, to be published in *Proceedings of NAG/DAGA*, Rotterdam (2009)

[36] B. C. J. Moore „Frequency Analysis and Masking“ in: B. C. J. Moore (ed.) *Hearing*, Academic Press, San Diego/New York, pp. 161-205 (1995)

[37] S. M. Salomons *Coloration and Binaural Decoloration of Sound Due to Reflections* doct. diss., Delft University of Technology (1995)

[38] A. Lindau, H.-J. Maempel, S. Weinzierl „Minimum BRIR grid resolution for dynamic binaural synthesis” In: *Proc. of the Acoustics '08*. Paris, pp. 3851-3856 (2008)

[39] ITU-R Rec. BS.1116-1 *Methods for the subjective Assessment of small Impairments in Audio Systems*

including Multichannel Sound Systems, Geneva, International Telecommunication Union (1997)

[40] ITU-R Rec. BS.1770 *Algorithms to measure audio programme loudness and true-peak audio level*, Geneva: International Telecommunication Union (2006)

[41] J. Bortz *Statistik für Human- und Sozialwissenschaftler*, Springer, Berlin/Heidelberg (2005)

[42] S. G. Norcross, M. Bochard „Inverse filtering design using a minimal-phase target function from regularizaion“ *Proc. of the 121st AES Convention*, San Francisco, preprint 6929 (October 2006)