Perceptual Evaluation of Headphone Compensation in Binaural Synthesis Based on Non-Individual Recordings

ALEXANDER LINDAU,[1] *AES Student Member,* **AND FABIAN BRINKMANN**[1]

(alexander.lindau@tu-berlin.de)　　　　　　　　(fabian.brinkmann@tu-berlin.de)

[1]*Audio Communication Group, Technical University of Berlin, Germany*

The headphone transfer function (HpTF) is a major source of spectral coloration observable in binaural synthesis. Filters for frequency response compensation can be derived from measured HpTFs. Therefore, we developed a method for measuring HpTFs reliably at the blocked ear canal. Subsequently, we compared non-individual dynamic binaural simulations based on recordings from a head and torso simulator (HATS) directly to reality, assessing the effect of non-individual, generic, and individual headphone compensation in listening tests. Additionally, we tested improvements of the regularization scheme of an LMS inversion algorithm, the effect of minimum phase inverse filters, and the reproduction of low frequencies by a subwoofer. Results suggest that while using non-individual binaural recordings the HpTF of the individual used for the recordings – typically a HATS – should be used for headphone compensation.

0 INTRODUCTION

0.1. Motivation

Binaural reproduction can achieve a high degree of realism. However, when directly comparing dynamic binaural synthesis to the corresponding real sound field we identified spectral coloration as a major shortcoming [1]. In this respect, the common practice to use head and torso simulators (HATS) for creating non-individual binaural recordings is especially problematic. Due to morphological differences, the head related transfer functions (HRTFs) differ from those of the actual listener and result in various distortions of auditory perception as described, for instance, by Møller et al. ([2]–[4]). Additionally, transducers involved in the binaural recording and reproduction signal chain introduce unwanted spectral coloration. These transducers include loudspeakers and microphones used for binaural measurements and the headphones used for reproduction. The influence of the headphone transfer function (HpTF) can potentially be compensated by inverse filtering. In an earlier study [5], comparing several inversion approaches for HpTFs, we found high-pass-regularized least-mean-square (LMS) inversion (cf. Kirkeby and Nelson [6]) approximating a pre-defined band pass as target function to be a perceptually well-suited inversion algorithm. However, coloration was still audible in these listening tests presumably originating both from using non-individual binaural recordings obtained with our HATS FABIAN [1] *and* from

using non-individual HpTFs for headphone compensation. As an approach to further optimize the headphone compensation in the case of non-individual binaural synthesis, in the present study we examined the effect of using non-individual, generic or individual HpTFs for headphone compensation.

0.2. State of the Art

Møller [7] has stated that all spatial information of the sound field is encoded in the sound pressure at the entrance of the blocked ear canal. In turn, the eardrum signal should be perfectly reproducible from the sound pressure measured at the blocked ear canal as long as headphones used for reproduction exhibit a linear frequency response at the blocked ear canal and an acoustic impedance close to that of free air (free air equivalent coupling, FEC [7]).

To make things difficult, different frequency response target functions deviating considerably from linearity have been defined for headphones [9]–[11]. Kulkarni and Colburn [13] showed that differences can be of the same order as found within HRTFs of different directions of incidence. Moreover, frequency response targets are approached most differently across manufacturers, models, and even within batches [5], [8], [12].

For circum- or extraaural headphones the situation is even more complicated: For the same headphone model, the individually differing morphology of the outer ear can

cause deviations up to 20 dB between individual HpTFs (inter-individual variability [8], [9]).

Møller et al. [8] found the inter-individual HpTF variability to be reduced when measuring at the blocked ear canal as compared to measuring at the open ear canal.

Transfer functions also vary as headphones are taken on and off repeatedly (intra-individual variability [5], [8], [12]). Therefore, Kulkarni and Colburn [13] recommended compensating the HpTF based on an average of multiple measurements taken while re-seating the headphones in-between. The authors further assumed leakage to be the dominating cause for intra-individual low-frequency variability observed with re-seating.

Assessing four different headphones in a criterion-free listening test Paquier and Koehl [14] found that positional variability leads to audible deviations.

Wenzel et al. [15] assessed the localization performance achievable with non-individual binaural recordings. Headphones were compensated using the HpTF of the individual whose HRTFs had been used for auralization (i.e a "reference subject"). Authors stated that recordings would be reproduced the less "faithful" the less the test subjects' HpTFs resembled that of the reference subject. While this might hold true, it must be kept in mind that these were still the wrong (i.e. the non-individual) HRTFs that were reproduced more faithful.

The benefit of individual over non-individual headphone compensation while rendering individual HRTFs, was illustrated by Pralong and Carlile [16]. Comparing two subjects' HpTFs they found deviations of up to 10 dB in the region of 3–7 kHz, which would in turn be distorting the HRTFs if applied as non-individual headphone compensation. On the other hand, the authors showed that using individual HpTFs for compensation leads to an assumed perceptually transparent reproduction (mean difference within ±1dB) of both individual and non-individual binaural recordings.

Martens [17] assessed the benefit of generic headphone compensation (i.e. a compensation filter based on the average HpTF obtained from several individuals). From a model-based test design the author concluded generic headphone compensation to be sufficient for faithful binaural synthesis.

By means of auditory filter analysis we assessed the effect of using non-individual, generic or individual HpTFs for compensation [18]. Results (cf. section 1.2) suggested that the fidelity of the equalization increases with the amount of individualization used in headphone compensation. Whether this trend was beneficial for auralization with non-individual recordings remained to be studied.

0.3. Scope of the Study

Directly comparing a non-individual binaural simulation to the respective real sound source, we aimed at assessing the effect of non-individual, generic, and individual headphone compensation on the perceived difference. Additionally, in [5] subjects occasionally mentioned pre-ringing and high frequency flaws. Therefore, we also assessed the fidelity of binaural reproduction while using minimum phase

Fig. 1. Custom-built silicone earplug. Left: CAD model, right: inserted into a subject's ear (from [18]).

instead of unconstrained phase filters for compensation, and for several improvements of the high-pass regularization inversion scheme to better adapt to the high frequency characteristics of HpTFs. Further on, as a possible future means to extend binaural reproduction beyond the lower cut-off frequency of headphones, we assessed the binaural reproduction's realism when it was combined with a subwoofer reproducing the low frequency components (50–166 Hz).

1 METHODS

1.1. Measuring Individual HpTFs

In [18] we presented custom-built silicone earplugs flush-cast with miniature electret condenser microphones (Knowles FG 23329, ø 2.5 mm) for measuring individual HpTFs at the blocked ear canal. Inserts were fabricated in three sizes based on anthropometrical data supplied by the manufacturer PHONAK (cf. Fig. 1).

For validating these earplugs, transfer functions were measured on a silicon-made artificial ear with ear canal, while reinserting the measurement microphone after each measurement. For comparison, these measurements were also conducted using two different types of foam earplugs: the EAR II plug and a less obtrusive and softer UVEX "com4-fit" foam plug, both commonly reported in binaural literature.

Due to replacement, the foam plug measurements showed deviations up to ±10 dB and more, whereas deviations obtained with our silicon earplugs were negligible below 8 kHz; above this range deviations reached ±2 dB (cf. Fig. 2).

Further on we measured HpTFs of 2 female and 23 male subjects using STAX SR 202 headphones and our silicone earplugs [18]. Subjects had to reposition the headphones before each of ten measurements. The spectral variability of all 250 measurements is depicted in Fig. 3. Four characteristic frequency ranges could be identified. Below 200 Hz (region I), differences of ±3 dB can primarily be assigned to leakage effects. Up to 2 kHz (region II), differences are smaller than 1 dB. Above 2 kHz and up to 5 kHz (region
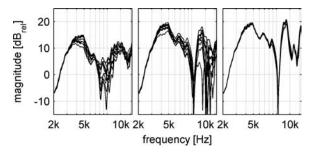
Fig. 2. Ten transfer functions of an artificial ear measured with Knowles FG 23329 and EAR II foam plug (left), UVEX foam plug (middle), and novel silicone earplug (right), when plugs were reinserted between each measurement (from [18]).
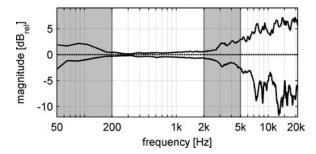

Fig. 3. Spectral variability of individual HpTFs (left ear only), upper and lower curve enclose the 75% percentile of the magnitude transfer functions. Shaded areas differentiate between characteristic regions of HpTF curves.
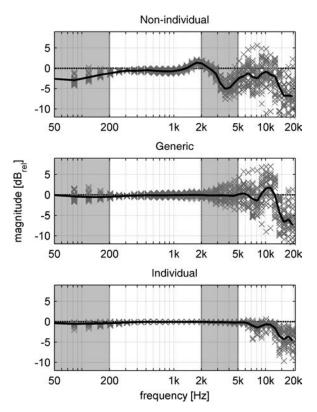

Fig. 4. Average differences of compensated HpTFs (both ears) of 25 subjects from target function for each band of an auditory filter bank and for three different inversion approaches. Grey crosses: average differences of singular subjects. Black solid curve: average difference across all subjects. LMS inversion with high-pass-regularization was used throughout (from [18]).

III), deviations quickly increase to ±3 dB. Above 5 kHz (region IV), the region of narrow pinna notches begins. Deviations in that range are distributed asymmetrically between approx. +7 and −11 dB, respectively.

1.2. Auditory Modeling of Inversion Results

We used the measured HpTFs to analyze non-individual, generic, and individual headphone compensation in a quantitative manner. Non-individual headphone compensation was realized by filtering each subject's HpTF with the inverse HpTF of a singular subject (here: HATS FABIAN). The inverse of the average HpTF across all 25 subjects served for the generic compensation, whereas for individual compensation the inverse of each subject's own average HpTF was applied. An auditory filter bank of 40 equivalent rectangular bandwidth (ERB) filters was used to model the perceptual deviation between compensated HpTFs and the target function comprising a band pass (−6 dB points: 50 Hz, and 21 kHz, 60 dB stop band rejection, cf. [5]).

When comparing compensation results to the overall HpTF variability (i.e. Fig. 4 to Fig. 3) it becomes clear, that the non-individual filter provides negligible improvement. As expected, the generic filter is found to symmetrically re-distribute the spectral deviations around the 0 dB line, while not reducing the overall amount of spectral variation. Individual compensation promises best results as regions I to III are nearly perfectly equalized.

Only the narrow pinna notches – typically occurring in HpTFs above 5 kHz – remain after compensation. The

preservation of notches is perceptually advantageous and directly intended when using the LMS inversion method with high-pass-regularization. For clarity, regularization means limiting the inversion effort in specific frequency regions. For regularization we used a shelve filter with 15 dB gain and a half-gain frequency of 4 kHz, resulting in a less precise compensation in the amplification region of the filter. As an adverse side effect of this type of regularization the lower plot in Fig. 4 reveals that high frequency damping in HpTFs is practically left uncorrected, potentially causing a dull or muffled reproduction. In connection with similar statements from subjects in [5], this was our motivation for assessing different improvements of the high-pass-regularization scheme in the present study.

1.3. Inverse HpTF Filter Design

Throughout this study raw HpTF measurements were shortened to 2048 samples; all inverse filters were designed to have the same length. Before inversion, the measurement microphones' frequency responses were removed from the HpTFs via deconvolution. As target function for the compensated headphones the above mentioned band pass was defined. The LMS method with high-pass-regularization allows designing HpTF inverse filters in the time [6] or in the frequency domain [19]. We used the latter method as it is faster, especially with larger filter lengths. With the
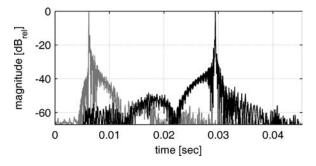
Fig. 5. Impulse responses of compensated HpTFs. Left/grey: using an LMS filter designed to exhibit minimum phase (acc. to [20]), right/black: using an LMS filter without constraints to filter phase designed to best approximate the (linear phase) target function after compensation.
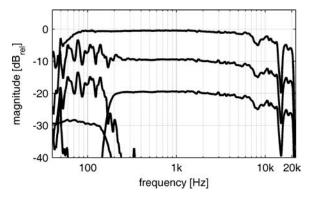

Fig. 6. Magnitude spectra of compensated HpTF (frequency domain LMS inversion with high-pass-regularization) measured at a test subject's right ear. Curves top down: 1) full range headphone reproduction, 2) sum response of 2–way reproduction, 3) 2–way reproduction, subwoofer and headphones shown separately, 4) near field response of subwoofer (all curves 1/24th octave smoothed, 10 dB offsets for clarity).

conventional LMS methods one typically defines the impulse response (the spectrum, resp.) of a linear phase band pass as target function. As a result the inverse filters may exhibit considerable pre-ringing (cf. Fig. 5). Lately, Norcross et al. [20] presented an approach to obtain inverse filters with minimum phase. We also tested this method in the listening test.

1.4. Subwoofer Integration

The STAX headphones could be equalized to reproduce at moderate levels a frequency range of 50–21 kHz (cf. Fig. 6, upper curve). Besides, as it might be of future interest to extend the reproduction to the full audio range, we tested integrating an active subwoofer into the binaural playback. Hereby it is assumed that binaural cues conveyed by headphone signals high pass filtered at some low frequency will still permit a proper perception of spatial sound. This is reasonable as ILD, ITD and spectral cues are nearly invariant over a larger frequency range (up to ca. 1 kHz). If we could show the "2-way"-reproduction's realism to be comparable to that of the headphones-only mode it might

render possible a low-frequency-extended reproduction of binaural content.

The ADAM SUB8 is a small (single 8'' driver) bass reflex design with adjustable gain and low pass cross over frequency. It can be fitted well beneath a listener's chair. For frequency response calibration near field measurements were conducted. Using two parametric equalizers from a loudspeaker controller (Behringer DCX2496) we could establish a nearly ideal 4th order band pass behavior within 26–150 Hz.

Fig. 6 shows measurements from the final calibration procedure in the listening test room collected at the ear canal entrance of a test subject while wearing the STAX headphones. Room modes disturbing the fidelity of the subwoofer reproduction were treated applying two more parametric equalizers of the DCX2496. A smooth summation of subwoofer and headphones output was achieved by level adjustment using the DCX2496 and phase delay adjustment using both a pre-delay in the target band pass – applied on the subwoofer to shape its lower slope – and the phase switch of the SUB8. After comparing calibration results from the test subject and the FABIAN HATS we assumed the alignment to be nearly invariant across subjects. In the listening test, the subwoofer was driven with a mono sum of the binaural headphone signals which were attenuated by 6 dB to preserve the level calibration.

This way, we were able to present the binaural signals in two different reproduction modes. In (a), the full range mode, only the headphones – equalized to approximate the target band pass response – were used, whereas in (b), the 2-way reproduction mode, headphone filters were designed to yield a crossover behavior at 166 Hz (-6 dB) reproducing the target band pass response in summation with the subwoofer.

2 LISTENING TEST I

Two listening tests were conducted. In the first we aimed at a perceptual evaluation of the three compensation approaches (non-individual, generic, individual). Conducting the second listening test was found necessary after inspecting the somewhat unexpected results of listening test I as will be explained in section 4.

In an acoustically dry recording studio ($RT_{1\,kHz} = 0.47$ s, $V = 235$ m^3), binaural room impulse responses (BRIRs) were measured using the FABIAN HATS. A measurement loudspeaker (Genelec 1030a) was placed frontally in a distance of 2 m, and BRIRs were measured for horizontal head movements within an angular range of $\pm\,80°$ and a step size of 1°. Our investigation was thus restricted to frontal sound incidence. In our opinion, for detecting spectral deficiencies of headphone compensation though, this familiar source setup made it most easy for subjects to focus on the task, thus resembling an – intended – worst case condition.

During measurements the HATS already wore the STAX headphones. These headphones are virtually transparent to exterior sound fields, in turn allowing simulation and reality to be directly compared without taking them off. Thus, by

applying dynamic auralization accounting for horizontal head movements [1] the virtual loudspeaker – presented via differently compensated headphones – could be directly compared to the real loudspeaker.

Besides (a) the three described approaches to headphone compensation (factor *filter*), we additionally assessed (b) type of content (pink noise and acoustic guitar, factor *content*), (c) the use of minimum phase versus unconstrained inverse filters (factor *phase*), and (d) the effect of a 2-way binaural reproduction with the low frequency content being reproduced by a calibrated subwoofer (factor *reproduction mode*) resulting in $3*2*2*2 = 24$ test conditions that were assessed in a fully repeated measures design (each subject assessed each condition). As we expected no interactions between tested factors, and while assuming an inter-subject correlation of 0.4, 20 subjects were calculated to be needed for testing a small main effect (E = 0.1) at a type-1 error level of 0.05 and a power of 80% [21], [22].

Subjects were seated in the former position of FABIAN in front of the real loudspeaker while their absolute position was controlled by aligning their ear canal entries using two perpendiculars. At the beginning of each listening test, the individual HpTFs were measured using our insert microphones and filters were calculated with prepared Matlab® routines. A training was conducted to familiarize subjects with stimuli and the rating process. In a multiple-stimulus ABC/HR listening test paradigm [23] 27 subjects (24 male, 3 female, average age 31.7 years) had to (a) detect the simulation and (b) rate its similarity to the real loudspeaker reproduction. On the graphical user interface, subjects found two sliders and three play buttons ("A", "B", "Ref/C") for each stimulus condition. The two buttons adjoining the sliders were randomly playing the test stimulus (HpTF-compensated simulation) or the reference (the real loudspeaker), the third button, "Ref/C" always reproduced the reference. Slider ends were labeled "identical" and "very different" (in German), and ratings were measured as continuous numerical values between 5 and 1. Only one of the two sliders could be moved from its initial position ("identical"), which would also indicate this sample as being identified as the test stimulus. While taking their time at will, subjects compared sub sets of six randomized stimuli using one panel of paired sliders. Within each sub set the audio content was kept constant. The length of the stimuli was about 5 seconds. For unbiased comparability the frequency response of the real loudspeaker was also limited by applying the target band pass. Additionally, real time indivualization of the interaural time delay [25] was used throughout the listening test. Including HpTF measurement, filter calculation, and training, the test took about 45–60 minutes per subject, of which on average 20 minutes were needed for rating.

3 RESULTS OF LISTENING TEST I

Results from two subjects were discarded after post-screening: one rated all simulations equally with "very different", another one experienced technical problems while testing. Following [23], results were calculated as differ-
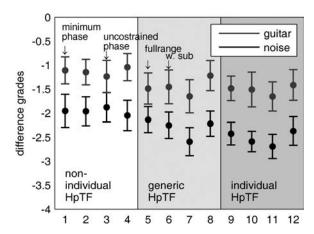

Fig. 7. Results from listening test I: Difference grades and 95% confidence intervals for all conditions averaged over all subjects. Shaded columns differentiate between filter types. Ratings for conditions *phase* and *reproduction mode* alternate throughout columns as indicated by arrows.

ence grades, subtracting the test stimulus' rating from the reference's rating. If the test stimulus was correctly identified all the time, only negative difference ratings would be observed (ranging from 0 = "identical" to −4 = "very different"). For all 24 test conditions average difference ratings and confidence intervals of the remaining 25 subjects are shown in Fig. 7.

Obviously, the simulation was always clearly detectable (negative average difference grades). This is not surprising as the ABC/HR design provides an open reference (i.e. the real loudspeaker is always played back when hitting the "Ref/C" button). Thus, slightest spectral deviation will enable subjects to rather easily detect the test stimulus, which in turn is likely the case as the binaural recordings were explained to be non-individual (cf. section 0.1).

The effect of *content* is also clearly obvious; moreover, for type of *filter* a noticeable variation can be seen. Effects of the conditions *phase* and *reproduction mode* are less obvious. As no intermediate anchor stimuli were defined, ratings were z-normalized across subjects before being subjected to inferential analysis (repeated measures ANOVA) [23]. In terms of average difference ratings we had formulated the following a-priori hypotheses for the four main effects: (a) $\mu_{individual} > \mu_{generic} > \mu_{non-indivdual}$, (b) $\mu_{guitar} > \mu_{noise}$, (c) $\mu_{minimum-phase} > \mu_{uncostrained-phase}$, (d) $\mu_{1-way} = \mu_{2-way}$. The inter-rater reliability was pleasingly high (Cronbach's α 0.944), indicating a sufficient duration of the training phase. We found effects for *content* and *filter* to be highly significant. In agreement with [5] and our a-priori hypothesis overall difference grades were significantly worse for the noise content. This is not surprising as the problematic frequency ranges of the compensated HpTF ranges (cf. Fig. 6) will be excited much stronger by wide band noise than with the rather limited frequency range of the guitar stimulus. The *filter* effect surprised us, as the simulation compensated with the non-individual HpTF (that of the FABIAN HATS) was rated best. Multiple comparisons

with Bonferroni adjustment furthermore showed that generic and individual compensation differed only insignificantly from each other, at least a trend was observed for the individual compensation to be rated worse. No significant effect of *phase* could be found, although there was a trend for unconstrained phase filters to be rated slightly worse. Additionally, and in accordance with our a-priori hypothesis, no effect of reproduction mode, i. e. no difference in the amount of perceived similarity with the real sound source could be found between headphones-only and 2-way reproduction mode. While – from post-hoc power calculation – having been able to reject a small effect size of E = 0.0924 with 80 % power, the latter null hypothesis can assumed to be well supported.

4 DISCUSSION OF RESULTS OF LISTENING TEST I

Although the 2-way reproduction showed moderate low frequency distortion (± 4 dB, Fig. 6), the amount of perceived similarity with the real sound source was of the same order as for full range headphone reproduction. Thus, results support the conclusion that with the application of moderate room equalization, proper level adjustment, and crossover design subwoofers might well be integrated into binaural reproduction for low frequency reproduction. Moreover, a future extension of the reproduction to the full audio range (i.e. down to 20 Hz) should be considered.

Regarding the effect of *filter*, from verbal responses of subjects we were already informed that when compared to reality, generic and individual compensation were perceived more damped in the high frequencies as compared to the non-individually compensated simulation. In order to understand what happened, we tried to reconstruct the signal differences subjects have perceived when comparing simulations and natural listening. Therefore, in the same setup as in listening test I, we measured five subjects' HpTFs and their BRIRs for frontal head orientation. Two different kinds of headphone compensation – (1) *non-individual* (HpTF from FABIAN), and (2) *individual* (HpTFs from each of the five subjects) – were applied to the subjects' HpTFs. Afterwards, HpTFs were convolved with FABIAN's frontal BRIR to obtain the signal our simulation would have produced at the five listeners' ears for a neutral head orientation. From comparinson of the spectrum of the subjects' own BRIRs and those of the differently compensated simulations (cf. Fig. 8 for spectral difference plots) we got an impression of the coloration people would have actually perceived in each of these situations.

While admitting that due to the small sample size this examination has an informal character, results confirmed that spectral differences (that were pronounced only above 5 kHz) were on average less in the case of non-individual headphone compensation. An explanation might be that the HpTF of FABIAN, measured with circumaural headphones, closely resembles a near-field HRTF preserving prominent spectral features from the pinna characterizing
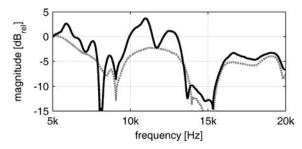

Fig. 8. Octave smoothed difference spectra of individual BRIRs and BRIRs from two binaural simulations using different headphone compensations (averaged across 5 subjects and both ears). Solid black curve: difference to non-individual BRIR compensated with non-individual HpTF, dashed grey curve: difference to non-individual BRIR compensated with individual HpTF.

also FABIAN's BRIRs used in the listening test. Using FABIAN's HpTF to compensate the headphone reproduction of FABIAN's binaural recordings may have resulted in a kind of de-individualization of the binaural simulation, especially compensating FABIAN's dominating high frequency (i.e. pinna-related) spectral characteristics. In contrast, when using the subjects own HpTF (individual compensation), the characteristics of the "foreign" BRIRs are reproduced nearly unaltered, meaning that inter-individual deviations become most audible.

We thus concluded that using HpTF of the subject that served also for non-individual binaural recordings was a special case not covered by our prior three-stage classification scheme of the *filter* types. To test our initial hypothesis again, we set up a new listening test, this time using a "true" non-individual HpTF, selected at random from the sample of listening test I (cf. section 6). Summing up, findings indicate that headphone compensation for binaural reproduction cannot be discussed without regarding the actual binaural recordings to be reproduced.

5 IMPROVING REGULARIZATION

As a new listening test was scheduled, we used the opportunity to test some more hypotheses. At first, we were concerned with improving the high-pass-regularization scheme. Two new methods were considered. The first is based on the assumption that a HpTF has to be compensated equally well within the complete pass band range (no general limitation in the high frequency range), while still taking care of 1–3 problematic notches typically occurring in HpTFs. A routine was programmed in Matlab®, which allowed us to define a regularization function that is flat on overall except for 1–3 parametric, peaking notch filters at the position of notches in the subject's HpTF. This in turn would limit the inversion effort only at the notches while flattening out all other deviations from linearity (termed "PEQ regularization" in the following). For the second approach, we assumed that regularization should somehow adapt to the HpTF, primarily flattening boosts while being of less effect with occurring notches. This behavior can be achieved by using the inverse average HpTF itself as a
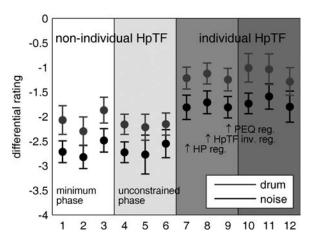
Fig. 9. Results from listening test II: Difference grades and 95% confidence intervals for all conditions averaged over all subjects. Lighter/darker shaded columns differentiate between filter types. Ratings for conditions *phase* and *regularization* alternate throughout columns as indicated by arrows.

regularization function [24]. We already tested this approach in [5] while using an octave smoothed version of the inverse HpTF. We considered inferior perceptual results in [5] to be due to the spectral resolution being too coarse. Therefore, this time we tested a sixth octave smoothed inverse HpTF (cf. [24]) as a regularization function (we termed this approach the "HpTF inverse regularization").

6 LISTENING TEST II

In the second listening test, we assessed effects of four factors: (a) the use of "individual" vs. "true non-individual" headphone compensation, with the latter being a HpTF related to *neither* the current test subject *nor* the binaural dataset in use (factor *filter*), (b) the two new regularization schemes (PEQ regularization, HpTF inverse regularization) and the high-pass-regularization (factor *regularization*), (c) again, the susceptibility to filter phase, this time using an assumed more critical stimulus, a drum set excerpt (factor *phase*), and (d) the type of content (pink noise, drum set, factor *content*). The listening test design was exactly the same as for test I. Again, the number of tested condition was 2*3*2*2 = 24. Maintaining all above mentioned specifications for test sensitivity and power, 27 new subjects (20 male, 7 female, average age 27.6 years) participated in the test.

7 RESULTS OF LISTENING TEST II

No subject had to be discarded in post-screening. The interrater reliability was again high (Cronbach's α 0.919). Average difference ratings and confidence intervals of the 27 subjects are shown in Fig. 9.

Overall detectability and the effect of *content* were comparable to test I. The effect of *filter* was now as expected. The "true" non-individual compensation was rated much worse than the individual condition. From comparison of

Fig. 7 and Fig. 9, true non-individual compensation can be assumed to be the worst choice in any of the tested cases. It though remains untested whether using no headphone compensation at all (cf. [5]) might be even worse. Effects of *phase* and *regularization* seem to be negligible. Standardized difference ratings were again subjected to repeated measures ANOVA. Effects for *content* and *filter* were found highly significant. Again, no susceptibility to filter *phase* (p = .98) could be found. Also, types of *regularization* showed no audible effect (p = .44), though there was a significant interaction (*filter*regularization*) indicating using the inverse smoothed HpTF for regularization to be best suited for individual HpTF compensation.

8 CONCLUSION

In two listening tests, we addressed the effect of different aspects of headphone compensation on the perceived difference of non-individual dynamic binaural synthesis when compared to reality. We assessed susceptibility to filter individualization, to filter phase, to audio content, the effect of a hybrid reproduction incorporating a subwoofer, and improvements of the high-pass-regularized LMS inversion scheme (the latter only for individual and "true" non-individual HpTF compensation). The effect of headphone compensation was found to be not straight forward. Surprisingly, non-individual binaural recordings that were headphone-compensated using the HpTF of the subject used for these recordings were perceived as most realistic. Due to the scope of this study, this conclusion remains limited to the case of non-individual recordings. With individual binaural recordings though, there is no reason why the individual HpTF should not be the best choice. A pronounced susceptibility to filter phase could not be found as well as an effect of two novel regularization schemes. A significant interaction though indicated the sixth octave smoothed inverse HpTF regularization to be more adequate in case of individual HpTF compensation. Using a cross over network, level, phase, and room correction calibrated at a reference subject's ear canal entrance, a subwoofer was shown suitable for low-frequency reproduction of binaural recordings.

9 ACKNOWLEDGMENTS

10 REFERENCES

[1] A. Lindau, T. Hohn, S. Weinzierl, "Binaural Resynthesis for Comparative Studies of Acoustical Environments", presented at the *122nd Convention of the Audio Engineering Society* (May 2007), convention preprint 7032.

[2] H. Møller et al. "Evaluation of Artificial Heads in Listening Tests", presented at the *102nd Convention of the Audio Engineering Society* (March 1997), convention preprint 4404.

[3] H. Møller et al. "Head-Related Transfer Functions of Human Subjects", In: *J. Audio Eng. Soc.*, vol. 43(5), pp. 300–332 (1995 May).

[4] H. Møller et al. "Binaural Technique: Do We Need Individual Recordings?", In: *J. Audio Eng. Soc.*, vol. 44(6), pp. 451–469 (1996 June).

[5] Z. Schärer, A. Lindau "Evaluation of Equalisation Methods for Binaural Signals" presented at the *126th Convention of the Audio Engineering Society* (May 2009), convention paper 7721.

[6] O. Kirkeby, P. A. Nelson, "Digital Filter Design for Inversion Problems in Sound Reproduction", In: *J. Audio Eng. Soc.*, vol. 47(7/8), pp. 583–595 (1999 July/August).

[7] H. Møller "Fundamentals of Binaural Technology" In: *Applied Acoustics*, 36(3/4), pp.171–218 (1992).

[8] H. Møller et al. "Transfer Characteristics of Headphones Measured on Human Ears", In: *J. Audio Eng. Soc.*, vol. 43(4), 203–217 (1995 April).

[9] J. R. Sank, "Improved Real-Ear Tests for Stereophones", In: *J. Audio Eng. Soc.*, vol. 28(4), 206–218 (1980 April).

[10] 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).

[11] H. Møller, "Design Criteria for Headphones", In: *J. Audio Eng. Soc.*, vol. 43(4), pp. 218–232 (1995 April).

[12] F. E. Toole, "The Acoustics and Psychoacoustics of Headphones" In: *Proc. of the 2nd Int. AES Conference: The Art and Technology of Recording* (1984 May), paper C1006.

[13] A. Kulkarni, H. S. Colburn, "Variability in the Characterization of the Headphone Transfer-Function", In: *J. Acoust. Soc. Am.*, 107(2), pp. 1071–1074 (2000).

[14] M. Paquier, V. Koehl, "Audibility of Headphone Positioning Variability", presented at the *128th Convention of the Audio Engineering Society* (2010 May), convention preprint 8147.

[15] E. M. Wenzel et al. "Localization Using Nonindividualized Head-Related Transfer Functions", In: *J. Acoust. Soc. Am.*, 94(1), pp. 111–123 (1993).

[16] D. Pralong, S. Carlile, "The Role of Individualized Headphone Calibration for the Generation of High Fidelity Virtual Auditory Space", In: *J. Acoust. Soc. Am.*, 100(6), pp. 3785–3793 (1996).

[17] W. L. Martens, "Individualized and Generalized Earphone Correction Filters for Spatial Sound Reproduction", In: *Proc. of ICAD 2003,*Boston.

[18] F. Brinkmann, A. Lindau "On the Effect of Individual Headphone Compensation in Binaural Synthesis", In: *Proc. of the 36th DAGA,* Berlin, pp. 1055–1056 (2010):

[19] O. Kirkeby et al. "Fast Deconvolution of Multichannel Systems Using Regularization", In: *IEEE Transactions on Speech and Audio Processing*, 6(2), pp. 189–195 (1998).

[20] S. G. Norcross, "Inverse Filtering Design Using a Minimal-Phase Target Function from Regularization", presented at the *121nd Convention of the Audio Engineering Society* (2006 October), convention preprint 6929.

[21] J. Bortz, N. Döring, *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*, 4. ed., pp. 618-630 (Heidelberg: Springer, 2006).

[22] F. Faul et al. "G*power 3: A Flexible Statistical Power Analysis Program for the Social, Behavioral, and Biomedical Sciences" In: *Behavior Research Methods*, 39(2), pp. 175–191 (2007).

[23] ITU-R Rec. BS.1116-1, *Methods for the Subjective Assessment of Small Impairments in Audio Systems including Multichannel Sound Systems* (Geneva 1997).

[24] S. G. Norcross; G. A. Soulodre, M. C. Lavoie "Evaluation of Inverse Filtering Techniques for Room/Speaker Equalization" presented at the *113th Convention of the Audio Engineering Society* (2002 October), convention paper 5662.

[25] A. Lindau, J. Estrella, S. Weinzierl, "Individualization of Dynamic Binaural Synthesis by Real Time Manipulation of the ITD" presented at *the 128th Convention of the Audio Engineering Society* (2010 May), convention paper 8088.

THE AUTHORS

Alexander Lindau

Fabian Brinkmann

Alexander Lindau was born 1976 in Berlin, Germany. In 2006 he received a M.A. degree in communication sciences, electrical engineering, and technical acoustics from the Technical University of Berlin, Germany. From 2007–2010 he held a Ph.D. grant of the Deutsche Telekom Laboratories working on the binaural synthesis of multichannel loudspeaker arrays. He is working on his Ph.D. thesis on the subjective comparison of real and reproduced concert hall experience using immersive audiovisual technologies. In 2011 he joined the research unit SEACEN (Simulation and Evaluation of Acoustical Environments), funded by the Deutsche Forschungsgemeinschaft. His fields of interest are room and electro acoustics, measurement techniques, binaural synthesis, auditory room simulation, and perceptual audio evaluation. Mr. Lindau is student member of AES, ASA, and DEGA (Germany).

●

Fabian Brinkmann was born 1983 in Bad Oeynhausen, Germany. He studied communication sciences and technical acoustics at the Technical University Berlin and recently finished his master thesis on "Individual Headphone Compensation for Binaural Synthesis." He is interested in spatial audio and signal processing. In 2011 he joined the DFG SEACEN research unit.