



Perceptual evaluation of individual headphone compensation in binaural synthesis based on non-individual recordings

Alexander Lindau¹, Fabian Brinkmann²

¹ Audio Communication Group, Technical University of Berlin, Germany

alexander.lindau@tu-berlin.de

Abstract

The headphone transfer function (HpTF) is one major source of spectral coloration that occurs in non-individual binaural synthesis. Filters for frequency response compensation can be derived from measurements of HpTFs. Therefore, a reliable method for measuring at the blocked ear canal had to be developed. Subsequently, in comparing dynamic binaural simulations directly to reality we could assess effects of non-individual, generic and individual headphone compensation in listening tests. Additionally, we tested improvements of the regularization scheme of a LMS inversion algorithm, the effect of minimum phase inverse filters, and a method for the low frequency extension by means of integrating a subwoofer.

Index Terms: binaural technique, individualization, headphone compensation

1. Introduction

1.1. Motivation

Binaural reproduction can achieve a high degree of perceptual plausibility. When directly comparing a binaural simulation to the corresponding real sound field, spectral coloration has been identified as a major shortcoming [1]. Different influences can be identified to be responsible for this coloration. Among these, using non-individualized binaural recordings (i.e. recordings made using a different subject than the listener) is one important aspect. This is typically the case if head and torso simulators (HATS) are used for binaural recordings. Due to morphologic differences the head related transfer functions (HRTFs) differ from those of the listeners resulting in various distortions of auditory perception [2], [3], [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. Whereas – in semi-diffuse sound fields – loudspeaker-introduced coloration can only be fully eliminated by imitating the frequency dependent directivity of the sound source to be simulated, for example through directional synthesis, the influence of the headphone transfer function (HpTF) can potentially be compensated by inverse filtering. According to a previous study comparing several inversion approaches for HpTFs [5], high-pass-regularized [1] least-mean-square (LMS) inversion [6] approximating a pre-defined band pass as target function proved to be a perceptively well-suited algorithm. However, in these listening tests – based on non-individual binaural recordings, too – audible spectral differences remained. As explained above, these differences originated both from using

non-individual binaural recordings obtained with our HATS FABIAN [1] and from using non-individual HpTFs (also measured on FABIAN) for headphone compensation. In the present study we examined the effect of using either non-individual, generic or individual HpTFs for headphone compensation in non-individual binaural synthesis while still applying the described LMS inversion algorithm. For clarity, it is emphasized here again that we did not consider the case of individual binaural recordings.

1.2. State of the art

In binaural technique it is generally assumed that the complete spatial information is included in the sound pressure at the eardrum or at the entrance of the blocked ear canal, respectively [7]. Accordingly, the eardrum signal should be perfectly reproducible from the sound pressure measured at the blocked ear canal, if headphones used for reproduction exhibit a) a linear frequency response at the blocked ear canal, and b) an acoustic impedance close to that of free air (free field equivalent coupling, FEC [7]). To make things difficult, different target functions deviating considerably from linearity have been defined for headphones' frequency responses [9] [10], [11]. This difference can be of the same order of magnitude as seen with HRTFs [13]. Moreover, frequency response targets are approached most differently across manufacturers, models and even within batches [5], [8], [12]. Therefore – and although there is a consensus that headphone compensation is needed in binaural synthesis [5], [8], [13] – a general compensation function for headphones cannot be defined. For circum- or extraaural headphones the situation is complicated further by the individually differing morphology of the outer ear. For the same headphone model, HpTFs may deviate up to 20 dB from individual to individual (inter-individual variability [8], [9]). Transfer functions also vary as headphones are taken on and off repeatedly (intra-individual variability [12], [5], [8]). Therefore, in [13] it was recommended to always compensate the HpTF based on an average of measurements taken while reseating the headphones multiple times. Also in [13], leakage was supposed as dominating cause for intra-individual low-frequency variability observed with reseating. With higher frequencies the sound field transmission in the cavity of typically used circumaural headphones becomes even more complicated and in turn more susceptible to reseating variability. In [8] and [14], HpTF variability was found to be less pronounced when measuring at the blocked than at the open ear canal. Assessing four different headphones in a criterion-free listening test in [15] it was shown that positional variability leads to audible deviation. Hence, intra-individual variability will always limit the performance of headphone equalization when using a non-adaptive filter. In [16], authors 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. Authors assumed that recordings would be reproduced the more “faithful” the more the test subjects’ HpTFs resembled that of the so-called ‘reference’ subject. Although the latter statement is true, it solely means, that it were the wrong (i.e. non-individual) HRTFs which were reproduced more faithful. The benefit of individual over non-individualized headphone compensation was illustrated in [17]. A comparison of two subject’s HRTFs and HpTFs showed deviations of up to 10 dB in the region of 3 - 7 kHz, which can be expected to affect non-individual headphone compensation. Additionally, using individual HpTFs for compensation was shown to satisfy proper (within < 1dB deviation) reproduction of both individual and non-individual binaural recordings at the listeners ear. The benefit of generic headphone compensation (i.e. a filter based on the average HpTF obtained from several individuals) was tried to be assessed in [18]. It is though assumed here, that the model-based test design in [18] was invalid, and that presented data did not to support the author’s opinion of generic headphone compensation being sufficient for faithful binaural synthesis. In [19], using a large sample of subjects, we assessed the effect of using non-individual, generic or individual HpTFs for compensation by means of auditory filter analysis. Results showed that a clear advantage of increasing individuality in headphone compensation can be expected.

1.3. Scope of the study

In a previous study [5], we found extraaural headphones to show a minimal intra-individual variability; additionally they are best suited for meeting the FEC criterion [8]. However, in the present study we restricted our investigation to the widely used circumaural STAX SR 2050 II headphones. The primary aim of the study was to support results from [19] with a formal listening test. In direct comparison to a real sound field we assessed the effect of non-individual, generic, and individual headphone compensation on the perceived authenticity of non-individual binaural recordings.

Additionally, in [5] subjects occasionally mentioned a) lacking low frequency (i.e. impact sound) fidelity, b) pre-ringing (especially with a musical stimulus), and c) high frequency deviation from the real sound field. We therefore also assessed a) the fidelity of binaural reproduction when it was extended with a subwoofer to reproduce the low frequency components (50-166 Hz), b) the use of minimum phase instead of linear phase filters inverse filters for compensation, and c) several improvements of the high-pass regularization inversion scheme to better fit the high frequency behavior of typical HpTFs.

2. Measuring individual HpTFs

At first, the need for headphone linearization necessitated a method for measuring individual HpTFs at the blocked ear canal. Consequently, in [19], we presented custom-built silicone earplugs flush-cast with miniature electret condenser microphones (Knowles FG 23329). Validation measurements showed that deviations due to replacing the earplugs were negligible below 8 kHz; above, they reached a maximum deviation of ± 2 dB. When compared to other reported methods using foam inserts, the new measurement method provides an increased reliability while being easy to conduct [19].

3. Auditory modeling of inversion results

Also in [19], HpTFs of two female and 23 male subjects were measured using the STAX 2050 II headphones. Measurements were repeated ten times per subject while reseating the headphones each time. Non-individual, generic, and individual HpTFs were constructed for headphone compensation. Non-individual headphone compensation was achieved by filtering with the inverse obtained from the HpTF of a singular subject (here: HATS FABIAN). The inverse of the average HpTF across all 25 subjects served for a 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 (-6dB points: 50 Hz to 21 kHz, 60 dB stop band rejection, see also [5]).

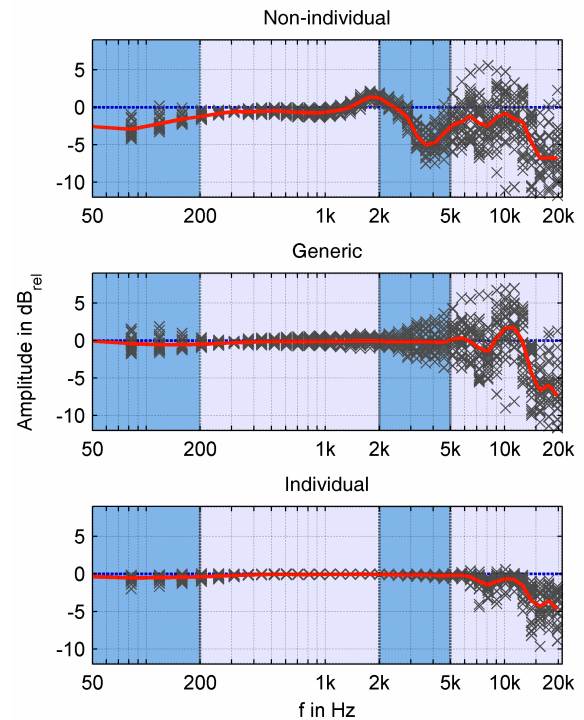


Figure 1: Average deviations of compensated HpTFs of 25 subjects from target function for each band of an auditory filter bank and for three different inversion approaches. Grey crosses: average deviation of single subjects. Red curve: average deviation across all subjects. Colored/shaded areas differentiate between characteristic features of HpTF curves. LMS inversion with high-pass-regularization was used throughout (from [19]).

From simulation results (Figure 1) it can be seen, that individual compensation promises best results. Only some narrow notches – typically occurring in HpTFs above 5 kHz – remain after compensation. The preservation of notches is a special advantage of the LMS inversion method with high-pass-regularization used here. Regularization can be understood as a frequency-dependent limitation of the inversion effort being proportionate to the regularization function’s gain. For high-pass-regularization we use a shelf filter with 15 dB gain and a half-gain frequency of 4 kHz. The lower plot in Figure 1 also reveals that diminishing high frequency gain of the headphones is left uncorrected

potentially causing a perception of high frequency damping. Therefore, we also assessed different improvements of the high-pass-regularization scheme.

4. Methods

4.1. Inverse HpTF filter design

Raw HpTF measurements were shortened to 2048 samples; all inverse filters were designed to have the same length. Before, 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 [20]. As the latter method is much faster, especially with increasing filter lengths, it was used throughout this study. With the conventional LMS methods one typically defines the impulse response (or the spectrum) of an ideal band pass as target function to be approximated. As the LMS error criterion is blind to whether remaining artifacts occur before or after the impulse's start, inverse filters and in turn the compensation results can exhibit considerable pre-ringing (cf. Figure 2). Lately, in [21], an approach was presented to obtain inverse filters with minimum phase. We therefore tested this method as an alternative to the conventional approach in the listening test.

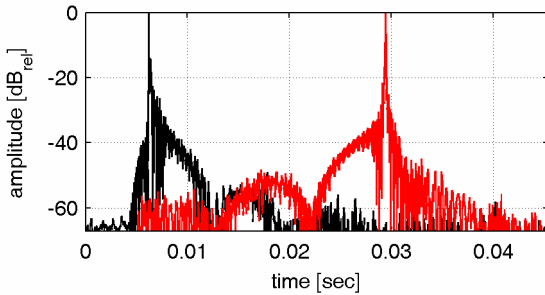


Figure 2: Impulse responses of compensated HpTFs. Left/black: compensated using an LMS filter designed to exhibit minimum phase (acc. to [21]), right/red: compensated using LMS filter designed without constraints to filter phase.

4.2. Subwoofer integration

The STAX headphones could be equalized to reproduce at moderate levels a frequency range of 50 - 20 kHz (cf. Figure 3, upper curve). In future application of binaural reproduction it might be of interest to circumvent this restriction and to extend the reproduction to the full audio range. We therefore tested integrating a commercially available active subwoofer into binaural playback. 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 an additional loudspeaker controller (Behringer DCX2496) we could established a nearly ideal a 4th order band pass behavior within 26 - 150 Hz (-3dB points). In the listening room, room modes disturb fidelity of low frequency reproduction. Therefore, additional room equalization (two more parametric EQs) had to be applied, again using the DCX2496. In 2-way reproduction mode, headphone filters were designed to be high-pass filtered at 166 Hz (-6dB) to – in summation with the subwoofer – reproduce

the complete audio range as defined by the target band pass. Figure 3 shows results from the calibration procedure measured at the ear canal entrance of one test subject (second author) while wearing the STAX headphones. This way we could easily adjust 2-way binaural reproduction mode to behave comparable to the headphone-only mode. A smooth summation of subwoofer and headphones was achieved by a) level adjustment using the controller and b) phase delay adjustment using both a pre-delay in the target band pass applied on the subwoofer and the phase switch of the SUB8. Comparing calibration results from the second author and the FABIAN robot, the alignment was found to be nearly invariant across subjects. In the listening test, the subwoofer was driven with a mono sum of the binaural signals. All stimuli were mono recordings. Thus, in order to maintain the correct level adjustment, the mono input signal had to be attenuated by 6 dB. For unbiased comparability to the headphone-only mode, the lower frequency response of the subwoofer was limited by additionally applying the target band pass. We did not extend the frequency response further downwards as for this time we were interested mainly in the question whether the 2-way approach was indistinguishable from the headphone-only mode. Once this perceptual equivalence has been assured, further extended usage is feasible, the more as we found the SUB8 capable of reproducing frequencies down to 26 Hz.

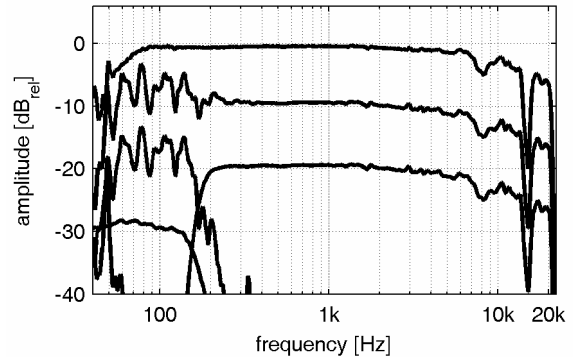


Figure 3: Magnitude spectra of compensated HpTF measured at second author's right ear, curves top down: 1) 1-way headphone reproduction, 2) sum response of 2-way reproduction, 3) 2-way reproduction, subwoofer and headphone shown separately, 4) near field response of subwoofer (all curves 1/24th octave smoothed, 10 dB offsets only for clarity)

5. 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). In an acoustically dry recording studio, 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 $\pm 80^\circ$ with a granularity of 1° . Using these datasets for dynamic auralization (fast time variant convolution accounting for horizontal rotational head movements of the listener via head tracking [1]) the virtual loudspeaker – presented via differently compensated headphones – could be directly compared to the real loudspeaker. During measurements the HATS already wore the STAX headphones. They are virtually transparent to exterior sound fields, allowing simulation and reality to be directly compared later on without taking off the headphones.

Besides a) the three described approaches to headphone compensation (factor *filter(3)*), we additionally assessed b) type of content (pink noise and acoustic guitar, factor *content(2)*), c) the use of minimum phase versus unconstrained phase inverse filters (factor *phase(2)*), and d) the effect of a 2-way binaural reproduction scenario with low frequency content being reproduced by a calibrated subwoofer (factor *reproduction mode(2)*) resulting in $3 \times 2 \times 2 \times 2 = 24$ test conditions in a fully repeated measures design. As we expected no interactions, assuming an intersubject 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% [22], [23].

Subjects were seated in the former position of FABIAN in front of the real loudspeaker. At the beginning of the listening test the individual HpTFs were measured and filters were calculated. Then, a training was conducted to familiarize subjects with stimuli and the rating process. In an ABC/HR listening test paradigm [24] 27 subjects (24 male, 3 female, avg. age 31.7 yrs.) had to detect and rate the simulation's similarity to the real loudspeaker. 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. Subjects could compare sub sets of six randomized stimuli using one panel of paired sliders (audio content kept constant within sub sets/panels) while taking their time at will. Stimuli were looped after 5 seconds. For unbiased comparability with the headphone simulation, the frequency response of the real loudspeaker was limited by applying the target band pass, too. Additionally, for achieving maximum possible authenticity of the dynamic binaural reproduction, individualization of ITD as described in [26] was used throughout the listening test. Including HpTF measurement, filter calculation, training and rating, the test took about 45-60 minutes per subject.

6. Results of listening test I

Two subjects were discarded in post-screening: one rated all simulation equally with 'very different', one experienced technical problems while testing. Following [24], results were calculated as difference grades, subtracting the test stimulus' rating from the true 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 Figure 4. Obviously, the simulation was always clearly detectable (negative difference grades). The effect of *content* is clearly obvious; moreover, for type of *filter* a noticeable variation can be seen. Effects of 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 (ANOVA) [24]. In terms of average difference ratings we had formulated the following a-priori hypotheses for the four main effects a) $\mu_{\text{individual}} > \mu_{\text{generic}} > \mu_{\text{non-individual}}$, b) $\mu_{\text{Guitar}} > \mu_{\text{Noise}}$, c) $\mu_{\text{Min-phase}} > \mu_{\text{Unconstrained-phase}}$, d) $\mu_{1\text{-way}} = \mu_{2\text{-way}}$. Based on

empirically found average intersubject correlation, post-hoc power analysis showed that we were able to test main effects at an effect size of $E = 0.0924$ at 0.05% type-I error level.

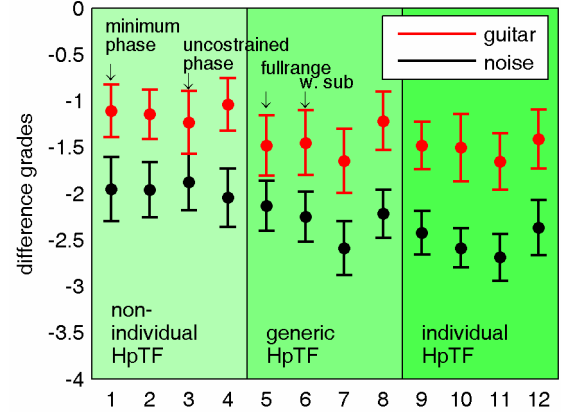


Figure 4: Results from listening test I: Difference grades and 95% CIs for all conditions averaged over all subjects. Colored/shaded columns differentiate between filter types. Ratings for conditions "phase" and "reproduction mode" alternate throughout columns as indicated by arrows.

The inter-rater reliability was pleasingly high (Cronbachs α 0.944), indicating a sufficient duration of the training phase. Despite for the 1st order interaction *filter*content* Mauchly's test of sphericity showed no violations of ANOVA preconditions. 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 HpTFs (cf. Figure 3) ranges will be excited much stronger by wide band noise than with the rather limited frequency range of the guitar stimulus. The effect of the applied compensation filter surprised us, as the simulation compensated with the non-individual HpTF (that of the FABIAN HATS) was rated best. Multiple comparisons (incl. Bonferroni adjustment) furthermore showed that generic and individual compensation differed only insignificantly from each other though a trend for the individual compensation to be rated worse was observed. No significant effect of filter phase could be found (tested one-sided acc. to a-priori hypothesis), though there was a trend for unconstrained phase filters to be rated slightly worse. Additionally, and in accordance with a-priori statements an effect of reproduction mode, thus discrimination between headphone-only and 2-way reproduction mode could not be found. The interpretation of this latter result is a special case, as here the null hypothesis (equality of means) was tested implicitly. Though, as a small effect size of $E = 0.0924$ could be rejected with 80% power this can be regarded a successful proof of the a-priori assumption.

7. Discussion of results of listening test I

From verbal responses we were already informed, that, when compared to reality, generic and individual compensation were perceived more damped in the high frequencies as the simulation compensated non-individually (using the FABIAN HpTF). In order to understand what happened, we reconstructed the signal difference between simulation and natural listening. Therefore, in the same setup as in listening test I, for five subjects we measured their HpTFs and BRIRs for frontal head orientation. One out of four different kinds of

compensation approaches: 1) *non-individual* (from FABIAN), 2) *individual* (from the five subjects), 3) *generic* (from the listening test I), and 4) using an arbitrary subject's HpTF (not FABIAN and not from one of the five subjects), thus a “true” *non-individual* compensation, was applied to the subjects' HpTFs. Afterwards, HpTFs were convolved with FABIAN's frontal BRIR to finally obtain the signal the simulation would have produced at the five listeners' ears. Comparing this result to the subjects' own BRIRs we got an impression of what difference people would have perceived in each of these four situations. Results confirmed that indeed – especially with higher frequencies – differences were minimal for the *non-individual* (FABIAN's HpTF) case. One explanation for that might be that the HpTF of FABIAN – as measured with a circumaural headphone – closely resembles a near-field HRTF, preserving prominent spectral features from the pinna which are contained also in binaural recordings conducted with this subject. Then, using the HpTF of the subject used also for the recordings may have resulted in a kind of de-individualization of the binaural simulation, especially compensating dominating individual spectral characteristics. This could explain why the simulation was perceived by an arbitrary subject to be more similar to listening with its own ears. 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. It is thus concluded that – at least in our case – using the subject's HpTF which served also for non-individual binaural recordings is a special case, which was not covered by the a-priori distinction between non-individual, generic, and individual headphone compensation. 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 9).

It was also interesting to note that though 2-way reproduction showed moderate low frequency variation (± 4 dB, Figure 3) there was nearly no audible difference. Thus, in aiming at equi-distributing low-frequency variability around the average reproduction level we succeeded in showing perceptual equality of both reproduction approaches.

8. Improving regularization

As a new listening test was scheduled, we wanted to use this opportunity to test some more hypotheses. The first is concerned with improving the high-pass-regularization scheme. As it was supposed to be causing additional high frequency loss, six new approaches to the regularization of inversion were implemented. After assessing their performance using the same auditory analysis as described in section 2, only two methods were considered further. They are described shortly in the following.

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 of HF-compensation), only taking care of 1-3 problematic notches typically occurring in HpTFs. A small tool was programmed in Matlab® allowing – based on the subjects average HpTF – to define a regularization function which is flat on overall except for 1-3 parametric, peaking notch filters at positions where the notches occurred in the HpTF. This in turn would limit inversion effort only at the notches while flattening out all other deviations from linearity (termed “PEQ regularization” in the following).

With the second approach it was assumed that regularization should somehow adapt to the HpTF, primarily flattening

boosts while being less aggressive when notches occur. This behavior can be achieved easily when using the inverse average HpTF itself as regularization function [25]. 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 this spectral resolution being too coarse. Therefore, this time we tested a sixth octave smoothed inverse HpTF (see also [25]) as regularization function (approach termed “HpTF inverse regularization”).

9. Listening test II

Eventually, in the second listening test, we assessed effects of four factors: a) the use of individual vs. “true” non-individual headphone compensation (factor *filter(2)*), b) the two new approaches to regularization (PEQ regularization, HpTF inverse regularization) in comparison to high pass regularization, (factor *regularization(3)*), c) again, the susceptibility to filter phase, this time using a thought more critical stimulus, a drum set excerpt (factor *phase(2)*), and d) the type of content (pink noise, drum set, factor *content(2)*). The listening test design was exactly the same as for test I. Again, the number of tested condition was $2 \times 3 \times 2 \times 2 = 24$. Maintaining all above mentioned specifications for test sensitivity and power, 27 subjects (20 male, 7 female, avg. age 27.6 yrs.) were acquired anew.

10. Results of listening test II

No subject had to be discarded in post-screening. From average intersubject correlation, post-hoc power analysis showed that this time we were able to test main effects at an effect size of $E = 0.0819$ at 0.05% type-I error level. The interrater reliability was high again (Cronbachs α 0.919). Average difference ratings and confidence intervals of the 27 subjects are shown in Figure 5.

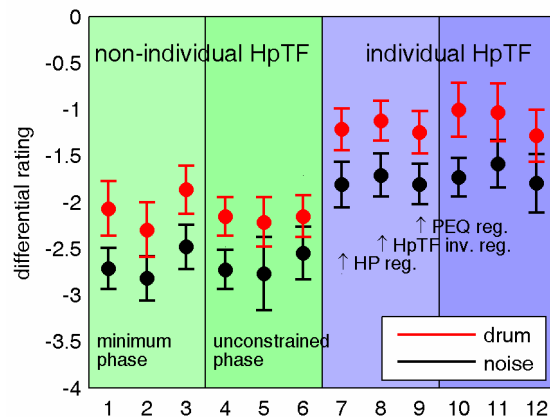


Figure 5: Results from listening test II: Difference grades and 95% CIs for all conditions averaged over all subjects. Greenish/bluish (lighter/darker) shaded columns differentiate between filter types. Ratings for conditions “phase” and “regularization” alternate throughout columns as indicated by arrows.

Again, the simulation was always clearly detectable. Also, the effect of *content* is clearly obvious again with noise being more critical. Using the true non-individual HpTFs the *filter* effect was now just as expected from auditory simulations, the (true) non-individual HpTF being rated much worse. From comparison of Figure 4 and Figure 5 true non-individual compensation can be assumed to be the worst choice in any

case. Subjects often described these simulations as being strongly colored and/or to exhibit audible ringing artifacts. Sometimes, even localization was said to be considerably impaired. With an individual HpTF these extreme artifacts of inverse filtering are avoided. Effects of *phase* and *regularization* seem to be negligible. Even for the drum set sample an advantage of the minimum phase design is not obvious. Z-normalized difference ratings were subjected to repeated measures ANOVA again. Mauchly's test of sphericity showed no violations of ANOVA preconditions. Effects for *content* and *filter* were proven to be highly significant. Again, no susceptibility to filter *phase* ($p=0.98$) could be found. Also, types of *regularization* showed no audible effect ($p=0.44$), though there was significant interaction (*filter*regularization*) indicating at least, that using the inverse smoothed HpTF for regularization to be best suited for individual HpTF compensation.

11. Conclusions

In two listening tests we compared the effect of different aspects of headphone compensation on the perceptual fidelity of non-individual dynamic binaural synthesis. 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 advantage of individualized headphone compensation was found to be not straight forward. Surprisingly, non-individual binaural recordings which were headphone-compensated using the HpTF of the subject used for these recordings were perceived most similar to reality. Even a generic compensation filter was – at least in trend – rated better than an individualized headphone compensation. Using an arbitrary subjects' HpTF though produced strongest audible artifacts and should be avoided in any circumstances. This conclusion is though limited to the case of non-individual recordings, and the described effect might have been pronounced by our listening test setup (frontal sound incidence, acoustically damped room). With individual binaural recordings though, there is no reason why the individual HpTF should not be the best choice. In this case, we recommend using LMS inversion with minimum phase inversion targets and a $1/6^{\text{th}}$ octave smoothed inverse of the subject's own HpTF as regularization function. However, a pronounced susceptibility to filter phase could not be found. Using constant level, phase, and room correction calibrated at a reference subject's ear canal entrance, a subwoofer was shown to be easily integrable for low-frequency extended headphone reproduction of binaural recordings.

12. Acknowledgements

Alexander Lindau was supported by a grant from the Deutsche Forschungsgemeinschaft (DFG, grant WE 4057/1-1).

13. References

- [1] Lindau, A.; Hohn, T. and Weinzierl, S.: "Binaural resynthesis for comparative studies of acoustical environments.", Proc. of the 122nd AES Conv., Vienna, preprint no. 7032, 2007
- [2] Møller, H. et al.: "Evaluation of Artificial Heads in Listening Tests." Proc. of the 102nd AES Conv., München, preprint no. 4404, 1997
- [3] Møller, H. et al.: "Head-Related Transfer Functions of Human Subjects." J. Audio Eng. Soc., Vol. 43, No. 5, pp. 300-332, 1995
- [4] Møller, H. et al.: "Binaural Technique: Do We Need Individual Recordings?" J. A. Eng. Soc., Vol. 44, No. 6, pp. 451-469, 1996
- [5] Schärer, Z. and Lindau, A.: "Evaluation of Equalisation Methods for Binaural Signals", Proc. of the 126th AES Conv., preprint 7721, 2009
- [6] Kirkeby, O. and Nelson, P. A.: "Digital Filter Design for Inversion Problems in Sound Reproduction.", J. Audio Eng. Soc., Vol. 47, No. 7/8, pp. 583-595, 1999
- [7] Møller, H.: "Fundamentals of Binaural Technology", Applied Acoustics, 36: 171-218, 1992
- [8] Møller, H. et al.: "Transfer Characteristics of Headphones Measured on Human Ears", J. Audio Eng. Soc., 43(4): 203-217, 1995
- [9] Sank, J. R.: "Improved Real-Ear Tests for Stereophones.", J. Audio Eng. Soc., 28(4), pp. 206-218, 1980
- [10] Theile, G.: "On the Standardization of the Frequency Response of High-Quality Studio Headphones.", J. Audio Eng. Soc., 34(12), pp. 956-969, 1986
- [11] Møller, H. et al.: "Design Criteria for Headphones.", J. Audio Eng. Soc., 43(4), pp. 218-232, 1995
- [12] Toole, F.E.: "The acoustics and psychoacoustics of headphones.", Proc. of the 2nd Int. AES Conference: The Art and Technology of Recording. Anaheim, CA., 1984
- [13] Kulkarni, A. and Colburn, H. S.: "Variability in the characterization of the headphone transfer-function.", J. Acoust. Soc. Am., 107(2), pp. 1071-1074, 2000
- [14] Riederer, K. A. J.: "Repeatability Analysis of Head-Related Transfer Function Measurements.", Proc. of the 105th AES Conv., San Francisco, preprint no. 4846, 1998
- [15] Paquier, M. and Koehl, V.: "Audibility of headphone positioning variability", Proc. of the 128th AES Conv., London, preprint no. 8147, 2010
- [16] Wenzel, E. M. et al.: "Localization using nonindividualized head-related transfer functions.", J. Acoust. Soc. Am., Vol. 94(1), pp. 111-123, 1993
- [17] Pralong, D. and Carlile, S.: "The role of individualized headphone calibration for the generation of high fidelity virtual auditory space.", J. Acoust. Soc. Am., 100(6), pp. 3785-3793, 1996
- [18] Martens, W. L.: "Individualized and generalized earphone correction filters for spatial sound reproduction.", Proc. of ICAD 2003 - 9th Meeting of the International Conference on Auditory Display. Boston, 2003
- [19] Brinkmann, F. and Lindau, A.: "The Effect of Individual Headphone Calibration in Dynamic Binaural Synthesis.", In: Proc. of the 36th DAGA. Berlin, pp. 1055-1056, 2010
- [20] Kirkeby, O. et al.: "Fast Deconvolution of Multichannel Systems Using Regularization." IEEE Transactions on Speech and Audio Processing, 6(2), pp. 189-195, 1998
- [21] Norcross, S.G. et al.: "Inverse Filtering Design Using a Minimal-Phase Target Function from Regularization", Proc. of the 121st AES Conv., preprint no: 6929, 2006
- [22] Bortz, J. and Döring, N.: Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler. 4. Aufl., Heidelberg: Springer, 2006
- [23] Faul, F. et al.: "G*power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences." Behavior Research Methods, 39(2), pp. 175-191, 2007
- [24] ITU (1997): ITU-R Rec. BS.1116-1: Methods for the subjective Assessment of small Impairments in Audio Systems including Multichannel Sound Systems, Geneva
- [25] Norcross, S.G.; Soulodre, G.A. and Lavoie, M.C.: "Evaluation of Inverse Filtering Techniques for Room/Speaker Equalization.", Proc. of the 113th AES Conv., Los Angeles, preprint no. 5662, 2002
- [26] Lindau, A.; Estrella, J. and Weinzierl, S.: "Individualization of dynamic binaural synthesis by real time manipulation of the ITD.", Proc. of the 128th AES Conv., London, preprint no. 8088, 2010