Proceedings of Meetings on Acoustics

Volume 19, 2013 http://acousticalsociety.org/

ICA 2013 Montreal
Montreal, Canada
2 - 7 June 2013

Psychological and Physiological Acoustics
Session 2aPPa: Binaural Hearing and Binaural Techniques II

2aPPa3. Physical correlates of loudness transfer functions in binaural synthesis

Florian Völk* and Hugo Fastl

***Corresponding author's address: AG Technische Akustik, MMK, Technische Universität München, Arcisstrasse 21, Munich, 80333, Bavaria, Germany, florian.voelk@mytum.de**

The frequency dependent level correction necessary for a binaural synthesis system to elicit via headphones the reference scene loudness of narrow-band signals is referred to as loudness transfer function. An ideal binaural synthesis system provides frequency independent loudness transfer functions for every listener. The frequency dependence of the average of a binaural synthesis system's individual loudness transfer functions has been shown to depend on the degree of individualization of the binaural synthesis system. In this contribution, perceptually acquired loudness transfer functions are compared from an auditory-adapted perspective to physical parameters of signals involved in the binaural synthesis process. The results provide quantitative relations between individual physical cues of the binaural synthesis output signals and the resulting loudness transfer functions.

Published by the Acoustical Society of America through the American Institute of Physics

Received 22 Jan 2013; published 2 Jun 2013

INTRODUCTION

Binaural synthesis (BS) with headphone (HP) playback is a physically motivated virtual acoustics technique intended to elicit the hearing sensations of a real or hypothetical reference scene by reproducing respectively synthesizing the corresponding sound pressure signals detected by the eardrums (cf. Møller, 1992; Völk, 2012b). The frequency dependent correction level necessary for a possibly suboptimal audio-transmission system to elicit the recording situation loudness of narrow-band signals is according to Völk and Fastl (2011) referred to as loudness transfer function (LTF). An ideal BS system provides frequency independent LTFs for every listener (Völk, 2010, 2011b). How close the ideal system is approximated by a BS implementation has been shown to depend on the implementation's degree of individualization, regarding loudness transfer (Völk and Fastl, 2011; Völk *et al.*, 2011a) as well as localization (Wenzel *et al.*, 1993; Møller *et al.*, 1996a,b,c; Kim and Choi, 2005).

In this contribution, the (perceptually acquired) LTFs discussed by Völk and Fastl (2011) are compared using auditory-adapted analysis methods (e. g. Terhardt, 1985; Völk *et al.*, 2011b) to physical parameters of signals involved in the BS process. The results provide quantitative relations between individual physical cues of the BS output signals and the resulting LTFs.

SYSTEM-THEORETIC BASICS

The BS system employed was implemented as described in Völk and Fastl (2011) with blocked auditory canal entrance recording based on the system-theoretical framework proposed by Völk (2011b). In the following, a graphical system overview is given based on Völk (2010) to the extent required for the definition of the signals and systems discussed in this paper.

Reference scene and recording situation

Let s_{ls} denote the digital sample sequence encoding the input signal to a loudspeaker (LS) box that represents the only sound source in the reference scene to be simulated by BS. The sound pressure time signals detected by an individual (*ind*) listener's eardrums in this reference scene are referred to as ear signals $\mathbf{p}_{\mathrm{e}}^{\mathrm{ind}}(\mathbf{x}_{\mathrm{h}}, \mathbf{x}_{\mathrm{ls}})$, which depend on the head and loudspeaker positions $\mathbf{x}_{\mathrm{h}}$ and $\mathbf{x}_{\mathrm{ls}}$. The upper branch in figure 1 represents the signal flow from s_{ls} to $\mathbf{p}_{\mathrm{e}}^{\mathrm{ind}}(\mathbf{x}_{\mathrm{h}}, \mathbf{x}_{\mathrm{ls}})$.

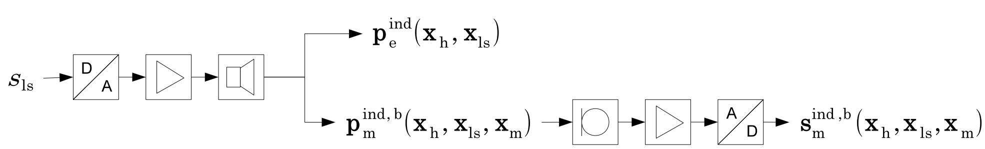

Figure 1: Block diagram of the blocked auditory canal binaural synthesis reference scene (top branch) and recording situation (bottom branch) discussed in this paper (figure modified from Völk, 2010). In the reference scene, a loudspeaker box driven by the digital sequence s_{ls} creates the ear signals $\mathbf{p}_{\mathrm{e}}^{\mathrm{ind}}(\mathbf{x}_{\mathrm{h}}, \mathbf{x}_{\mathrm{ls}})$. In the recording situation, the sound pressures $\mathbf{p}_{\mathrm{m}}^{\mathrm{ind,b}}(\mathbf{x}_{\mathrm{h}}, \mathbf{x}_{\mathrm{ls}}, \mathbf{x}_{\mathrm{m}})$ detected by miniature microphones at the blocked auditory canal entrances are encoded in the sequences $\mathbf{s}_{\mathrm{m}}^{\mathrm{ind,b}}(\mathbf{x}_{\mathrm{h}}, \mathbf{x}_{\mathrm{ls}}, \mathbf{x}_{\mathrm{m}})$.

Ideally, BS is implemented based on measuring the ear signals (the sound pressures actually detected by the eardrums), which is difficult for human ears since the pressure across the eardrum varies, especially at high frequencies (cf. Stinson, 1985; Hudde and Aumann, 2010).

The BS method discussed in this paper is therefore based on miniature microphone recording of the sound pressures $\mathbf{p}_{\mathrm{m}}^{\mathrm{ind,b}}(\mathbf{x}_{\mathrm{h}}, \mathbf{x}_{\mathrm{ls}}, \mathbf{x}_{\mathrm{m}})$ at the positions $\mathbf{x}_{\mathrm{m}}$ at the entrances to the blocked auditory canals. The microphones (Sennheiser KE 4-211-2 in amplifier configuration) were mounted according to Møller *et al.* (1995b) in foam ear plugs and inserted some millimeters in the ear canal, with the aim of recording at a position where the pressure field is independent of the sound incidence direction (Hammershøi and Møller, 1996; Hudde and Schmidt, 2009). After amplification and analog-digital-conversion (A/D; RME Fireface 400, 24 Bit, 44.1 kHz), the sequences $\mathbf{s}_{\mathrm{m}}^{\mathrm{ind,b}}(\mathbf{x}_{\mathrm{h}}, \mathbf{x}_{\mathrm{ls}}, \mathbf{x}_{\mathrm{m}})$ represent the sound pressures $\mathbf{p}_{\mathrm{m}}^{\mathrm{ind,b}}(\mathbf{x}_{\mathrm{h}}, \mathbf{x}_{\mathrm{ls}}, \mathbf{x}_{\mathrm{m}})$ detected by the miniature microphones at the blocked auditory canal entrances in the reference scene (lower branch in figure 1). In other words, the reference scene is encoded by the blocked auditory canal entrance sound pressures $\mathbf{p}_{\mathrm{m}}^{\mathrm{ind,b}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{m_{rec}}})$, knowing that the reference scene ear signals $\mathbf{p}_{\mathrm{e}}^{\mathrm{ind}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}})$ arise in this situation if the plugs are removed.

When actually implementing a BS system based on blocked auditory canal entrance measurement, typically the so-called recording situation transfer functions

$$\mathbf{H}_{\mathrm{rec_m}}^{\mathrm{ind,b}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{m_{rec}}}) = \frac{\mathbf{S}_{\mathrm{m}}^{\mathrm{ind,b}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{m_{rec}}})}{S_{\mathrm{ls}}} \tag{1}$$

are acquired in the reference scene, which relate the discrete Fourier spectra $\mathbf{S}_{\mathrm{m}}^{\mathrm{ind,b}}(\mathbf{x}_{\mathrm{h}}, \mathbf{x}_{\mathrm{ls}}, \mathbf{x}_{\mathrm{m}})$ and S_{ls} of the digital sequences representing the LS input and miniature microphone output sequences $\mathbf{s}_{\mathrm{m}}^{\mathrm{ind,b}}(\mathbf{x}_{\mathrm{h}}, \mathbf{x}_{\mathrm{ls}}, \mathbf{x}_{\mathrm{m}})$ and s_{ls} (cf. Völk, 2011b). The time domain representations of equation 1 are referred to as recording situation impulse responses $\mathbf{h}_{\mathrm{rec_m}}^{\mathrm{ind,b}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{m_{rec}}})$.

Playback and headphone transfer function measurement situation

In this paper, BS with HP playback is discussed. Let $\mathbf{s}_{\mathrm{hp}}$ denote the digital sequences encoding the HP input signals. The sound pressure time signals detected by the eardrums of an individual using the HPs in this playback situation are referred to as ear signals under the HPs $\mathbf{p}_{\mathrm{e}}^{\mathrm{ind,h}}(\mathbf{x}_{\mathrm{hp}})$, which depend on the positions $\mathbf{x}_{\mathrm{hp}}$ of the HPs on the ears (cf. Völk, 2011a). The upper branch in figure 2 represents the signal flow from $\mathbf{s}_{\mathrm{hp}}$ to $\mathbf{p}_{\mathrm{e}}^{\mathrm{ind,h}}(\mathbf{x}_{\mathrm{hp}})$.

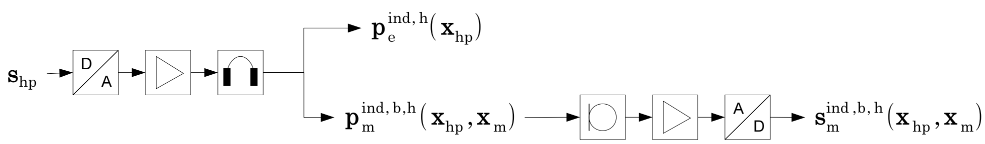

Figure 2: Block diagram of the binaural synthesis playback situation (top branch) and headphone transfer function measurement situation (bottom branch) discussed in this paper (figure modified from Völk, 2010). In the playback situation, headphones driven by the digital sequences $\mathbf{s}_{\mathrm{hp}}$ create the ear signals $\mathbf{p}_{\mathrm{e}}^{\mathrm{ind,h}}(\mathbf{x}_{\mathrm{hp}})$. During the headphone transfer function measurement, the sound pressures $\mathbf{p}_{\mathrm{m}}^{\mathrm{ind,b,h}}(\mathbf{x}_{\mathrm{h}}, \mathbf{x}_{\mathrm{ls}}, \mathbf{x}_{\mathrm{m}}, \mathbf{x}_{\mathrm{hp}})$ detected by miniature microphones at the blocked auditory canal entrances are encoded in the sequences $\mathbf{s}_{\mathrm{m}}^{\mathrm{ind,b,h}}(\mathbf{x}_{\mathrm{hp}}, \mathbf{x}_{\mathrm{m}})$.

The lower branch of figure 2 represents the sound pressures $\mathbf{p}_{\mathrm{m}}^{\mathrm{ind,b,h}}(\mathbf{x}_{\mathrm{hp}}, \mathbf{x}_{\mathrm{m}})$ detected by miniature microphones mounted according to Møller *et al.* (1995a) in foam ear plugs at the blocked auditory canal entrances, as well as their encoding in the digital sequences $\mathbf{s}_{\mathrm{m}}^{\mathrm{ind,b,h}}(\mathbf{x}_{\mathrm{hp}}, \mathbf{x}_{\mathrm{m}})$. The relations of the corresponding discrete Fourier spectra $\mathbf{S}_{\mathrm{m}}^{\mathrm{ind,b,h}}(\mathbf{x}_{\mathrm{hp}}, \mathbf{x}_{\mathrm{m}})$ to the respective HP driving spectra $\mathbf{S}_{\mathrm{hp}}$ are often referred to as headphone transfer functions (HPTFs)

$$\mathbf{H}_{\mathrm{hptf_m}}^{\mathrm{ind,h,b}}(\mathbf{x}_{\mathrm{hp_{hptf}}}, \mathbf{x}_{\mathrm{m_{hptf}}}) = \frac{\mathbf{S}_{\mathrm{m}}^{\mathrm{ind,b,h}}(\mathbf{x}_{\mathrm{hp_{hptf}}}, \mathbf{x}_{\mathrm{m_{hptf}}})}{\mathbf{S}_{\mathrm{hp}}}, \tag{2}$$

with corresponding headphone impulse responses $\mathbf{h}_{\mathrm{hptf_m}}^{\mathrm{ind,h,b}}(\mathbf{x}_{\mathrm{hp_{hptf}}}, \mathbf{x}_{\mathrm{m_{hptf}}})$. In the HPTF measurement situation characterized by $\mathbf{p}_{\mathrm{m}}^{\mathrm{ind,b,h}}(\mathbf{x}_{\mathrm{hp_{hptf}}}, \mathbf{x}_{\mathrm{m_{hptf}}})$, the ear signals $\mathbf{p}_{\mathrm{e}}^{\mathrm{ind,h}}(\mathbf{x}_{\mathrm{hp_{hptf}}})$ arise if the plugs are removed.

Stable and, considering the audible and technically available frequency and dynamic range, meaningfully designed equalization filters (*eq*) with the target transfer functions

$$\mathbf{H}_{\mathrm{eq_m}}^{\mathrm{ind,b}}(\mathbf{x}_{\mathrm{hp_{hptf}}}, \mathbf{x}_{\mathrm{m_{hptf}}}) = \frac{1}{\mathbf{H}_{\mathrm{hptf_m}}^{\mathrm{ind,h,b}}(\mathbf{x}_{\mathrm{hp_{hptf}}}, \mathbf{x}_{\mathrm{m_{hptf}}})} \tag{3}$$

must be applied for equalization purposes in the BS process (Völk, 2011b). The corresponding equalization filter impulse responses are denoted by $\mathbf{h}^{\mathrm{ind,b}}_{\mathrm{eq_m}}(\mathbf{x}_{\mathrm{hp_{hptf}}}, \mathbf{x}_{\mathrm{m_{hptf}}})$.

Binaural synthesis situation

In the actual BS situation, the HPs are driven by the signal encoded in the digital sample sequences $\mathbf{s}^{\mathrm{ind,b}}_{\mathrm{hp_{bs,m}}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{m_{rec}}}, \mathbf{x}_{\mathrm{hp_{hptf}}}, \mathbf{x}_{\mathrm{m_{hptf}}})$ corresponding to the discrete Fourier spectra

$$\begin{aligned}\mathbf{S}^{\mathrm{ind,b}}_{\mathrm{hp_{bs,m}}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{m_{rec}}}, \mathbf{x}_{\mathrm{hp_{hptf}}}, \mathbf{x}_{\mathrm{m_{hptf}}}) &= \\ = \mathbf{H}^{\mathrm{ind,b}}_{\mathrm{rec_m}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{m_{rec}}}) \cdot \mathbf{H}^{\mathrm{ind,b}}_{\mathrm{eq_m}}(\mathbf{x}_{\mathrm{hp_{hptf}}}, \mathbf{x}_{\mathrm{m_{hptf}}}) \cdot S_{\mathrm{ls}}.\end{aligned} \tag{4}$$

Figure 3 shows a block diagram of the BS situation, where the upper branch represents the part implemented by the recording situation transfer functions, the middle branch the equalization, and the lower branch the actual playback in the BS situation.

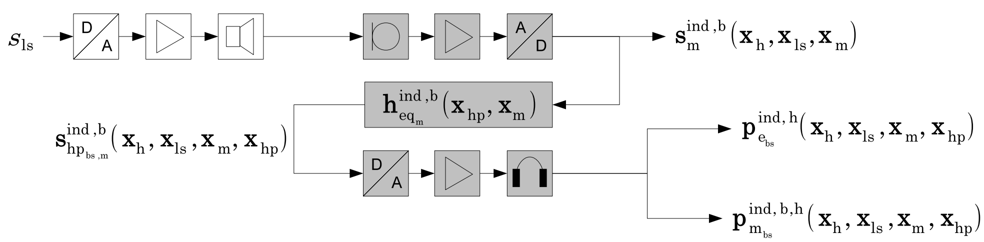

Figure 3: Block diagram of the full binaural synthesis procedure with blocked auditory canal entrance miniature microphone measurement discussed in this paper (figure modified from Völk, 2010). The upper branch represents the recording situation, the middle branch the headphone transfer functions-based equalization and the lower branch the playback situation (cf. figures 1 and 2).

Figure 3 indicates graphically that, idealistically assuming ideal equalization and all involved positions to be kept constant, the reference scene sound pressures detected by miniature microphones at the blocked auditory canal entrances can be reproduced by BS, formulated by

$$\mathbf{p}^{\mathrm{ind,b,h}}_{\mathrm{m_{bs}}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{m_{rec}}}, \mathbf{x}_{\mathrm{hp_{play}}}, \mathbf{x}_{\mathrm{hp_{hptf}}}, \mathbf{x}_{\mathrm{m_{hptf}}}) = \mathbf{p}^{\mathrm{ind,b}}_{\mathrm{m}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{m_{rec}}}), \tag{5}$$

which is proven mathematically by Völk (2011b). The subsystems mutually canceling each others influences are marked gray. However, the eventual goal of BS is reproducing the reference scene ear signals, that is

$$\mathbf{p}^{\mathrm{ind,h}}_{\mathrm{e_{bs}}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{m_{rec}}}, \mathbf{x}_{\mathrm{hp_{play}}}) \stackrel{!}{=} \mathbf{p}^{\mathrm{ind}}_{\mathrm{e}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}). \tag{6}$$

Assuming the validity of equation 5, the binaurally synthesized equal the reference scene ear signals if the ear signals resulting for given sound pressures at the blocked auditory canal entrance microphones when the plugs are removed are identical in the loudspeaker and headphone playback situations. The differences between those situations are the physical presence of the headphones and the different sound sources. It can be shown that the validity of equation 6 given equation 5 depends on the headphone model used (cf. Møller *et al.*, 1995a; Völk, 2012a).

INDIVIDUALIZATION ASPECTS

Recording (equation 1), headphone transfer function measurement (equation 2), and consequently the equalization (equation 3) for BS must be implemented individually for reaching the physically perfect synthesis. Due to the high effort of implementing fully individual BS, especially if a dynamic system adapting to listener movements is to be built, different approaches of reducing the complexity are commonly used.

Nonindividual recording

The recording for BS (equation 1) is often carried out nonindividually, that is with a subject different from the actual listener (Wenzel *et al.*, 1991; Bronkhorst, 1995; Møller *et al.*, 1996c) or using an artificial head (Møller *et al.*, 1997, 1999).

Nonindividual equalization

Regarding the equalization filter design, it is possible to take into account only the magnitude spectrum of the equalization target given by equation 3. In contrast to fully individualized equalization, this procedure is referred to as individual magnitude equalization (cf. Völk and Fastl, 2011). Nonindividual magnitude equalization is defined in an analogous manner based on a nonindividual equalization filter target, derived from headphone impulse responses (HPIRs) measured on a subject different from the listener (Møller *et al.*, 1995a; Völk, 2011a) or an artificial head (Kulkarni and Colburn, 2000).

The magnitude spectra of human outer ear transfer functions show extreme values at frequencies typical for the specific subject (Mehrgardt and Mellert, 1977). Consequently, averaging different individual equalization targets results in a loss of individual characteristics. However, if the magnitude spectrum average is computed with the aim of providing a BS equalization target improving the synthesis for arbitrary subjects, it is intended to include only the transfer characteristics common to most subjects. Combined with a linear phase response, these filters are referred to as average magnitude equalization filters (Völk and Fastl, 2011).

PROCEDURE AND HYPOTHESES

Since this paper attempts establishing relationships between individual (physical) stimulus cues and the resulting hearing sensations, the actual binaurally synthesized ear signals $\mathbf{p}^{\mathrm{ind,h}}_{\mathrm{e_{bs}}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{m_{rec}}}, \mathbf{x}_{\mathrm{hp_{play}}})$ were the ideal physical signals to be compared. However, ear signal definition and measurement are complicated, because the pressure across the eardrum varies, especially at high frequencies (cf. Stinson, 1985; Hudde and Aumann, 2010). For that reason, a different approach is selected here in that the HP input signals $\mathbf{s}^{\mathrm{ind,b}}_{\mathrm{hp_{bs,m}}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{m_{rec}}}, \mathbf{x}_{\mathrm{hp_{hptf}}}, \mathbf{x}_{\mathrm{m_{hptf}}})$ respectively the related transfer functions

$$\begin{aligned}\mathbf{H}^{\mathrm{ind,b}}_{\mathrm{hp_{in,bs}}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{m_{rec}}}, \mathbf{x}_{\mathrm{hp_{hptf}}}, \mathbf{x}_{\mathrm{m_{hptf}}}) &= \frac{\mathbf{S}^{\mathrm{ind,b}}_{\mathrm{hp_{bs,m}}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{m_{rec}}}, \mathbf{x}_{\mathrm{hp_{hptf}}}, \mathbf{x}_{\mathrm{m_{hptf}}})}{S_{\mathrm{ls}}} \\ &= \mathbf{H}^{\mathrm{ind,b}}_{\mathrm{rec_m}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{m_{rec}}}) \cdot \mathbf{H}^{\mathrm{ind,b}}_{\mathrm{eq_m}}(\mathbf{x}_{\mathrm{hp_{hptf}}}, \mathbf{x}_{\mathrm{m_{hptf}}})\end{aligned} \tag{7}$$

derived from equation 4 are discussed. These transfer functions are compared for different degrees of individualization to the corresponding LTFs, that is the input level difference $L_{\mathrm{bs_{in}}} - L_{\mathrm{ls_{in}}}$ between the binaurally synthesized and the real LS at equal loudness for pure tones. This procedure is motivated and discussed in the following.

The situation described by equation 7 deviates from the overall BS situation in that the necessarily individual playback situation (lower branch of figure 3) is not included. However, the parts of the overall BS situation that vary with the degree of individualization are included in both cases. Consequently, in completely nonindividualized situations, the transfer functions $\mathbf{H}^{\mathrm{nind,b}}_{\mathrm{hp_{in,bs}}}$ as defined by equation 7 are identical for every subject, while in a fully individualized system, different transfer functions $\mathbf{H}^{\mathrm{ind,b}}_{\mathrm{hp_{in,bs}}}$ occur for each subject.

Figure 4 shows median (black) and inter-quartile range (gray) of the LTFs given by Völk and Fastl (2011) for individually equalized BS with individual recording. The BS system was implemented with miniature microphone measurement at the blocked auditory canals entrances.

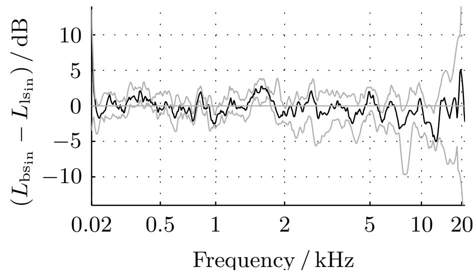

Figure 4: Median (black) and inter-quartile range (gray) of the individual level differences between the input signals to a binaurally synthesized and the corresponding real loudspeaker in front of the subjects in a reverberant laboratory, adjusted to equal loudness by Békésy-Tracking. Headphone specimen Sennheiser HD 800, static synthesis, individual recording, individual magnitude and phase equalization. Data from Völk and Fastl (2011).

Figure 4 reveals that individually equalized BS with individual recording using miniature microphones at the entrances of the blocked auditory canals provides for a frontally located loudspeaker on average frequency independent LTFs with the accuracy of the psychoacoustic tracking procedure employed. This scenario represents the best case result of the situations studied by Völk and Fastl (2011). For that reason, the corresponding individual transfer functions defined by equation 7 are considered the reference transfer functions $\mathbf{H}^{\mathrm{ind,b}}_{\mathrm{hp_{in,bs,ref}}}$, providing on average nearly frequency independent LTFs.

Völk (2013) – based on Theile (1980) – puts forward the hypothesis that the human hearing system removes the individual spectro-temporal cue patterns evaluated in determining the hearing sensation position from the physical inputs before forming a loudness percept. Consequently, cues not evaluated in the localization process would be passed on to the loudness perception stage, influencing the LTF and therefore influencing what is often referred to as sound color.

Assuming further $\mathbf{H}^{\mathrm{ind,b}}_{\mathrm{hp_{in,bs,ref}}}$ to provide the individually optimal headphone input signals resulting in the individually optimal ear signals, all deviating $\mathbf{H}^{\mathrm{ind,b}}_{\mathrm{hp_{in,bs}}}$ would result in frequency dependent LTFs, especially for deviating magnitude spectra. Based on these hypotheses, the frequency dependent level difference

$$L_{\mathrm{ref}} - L = 20\log_{10}\frac{\left|\mathbf{H}^{\mathrm{ind,b}}_{\mathrm{hp_{in,bs,ref}}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{m_{rec}}}, \mathbf{x}_{\mathrm{hp_{hptf}}}, \mathbf{x}_{\mathrm{m_{hptf}}})\right|}{\left|\mathbf{H}^{\mathrm{ind,b}}_{\mathrm{hp_{in,bs}}}(\mathbf{x}_{\mathrm{h_{ref}}}, \mathbf{x}_{\mathrm{ls_{ref}}}, \mathbf{x}_{\mathrm{m_{rec}}}, \mathbf{x}_{\mathrm{hp_{hptf}}}, \mathbf{x}_{\mathrm{m_{hptf}}})\right|}\,\mathrm{dB} \tag{8}$$

provides a prognosis of the frequency dependence of the individual LTF to be expected based on the nonindividual magnitude spectra applied. The validity of this prognosis and of the corresponding hypotheses is discussed in the following.

For the analysis, the transfer functions were computed according to Völk (2013) by auditory-adapted analysis (AAA) of the 256 impulse response samples centered around the impulse response maximum. The variability in each resulting set S of auditory-adapted transfer characteristics (that is magnitude spectrum l_κ and group delay τ_{g_κ}, cf. Völk, 2013) is in accordance to Völk (2011a) addressed using the quartiles, the 25 %, 50 %, and 75 % percentiles $\mathrm{P}_{25}(S)$, $\mathrm{P}_{50}(S)$, and $\mathrm{P}_{75}(S)$ at each analysis frequency f_{A_κ} separately for l_κ and τ_{g_κ}. In this paper, only the magnitude spectra according to equation 8 are discussed.

RESULTS AND DISCUSSION

The results presented in this paper are shown as two-column figures, with the frequency dependent medians (black) and inter-quartile ranges (gray) of the individual LTFs $L_{\mathrm{bs_{in}}} - L_{\mathrm{ls_{in}}}$ given by Völk and Fastl (2011) in the left subfigure and the corresponding inter-individual medians (black) and inter-quartile ranges (gray) of the frequency dependent level difference $L_{\mathrm{ref}} - L$ according to equation 8 in the right subfigure. Shown are the results of ten subjects, the BS was implemented using a specimen of Sennheiser HD 800 headphones. Gray horizontal bars on the abscissae of the left panels indicate median deviations exceeding the methodical accuracy.

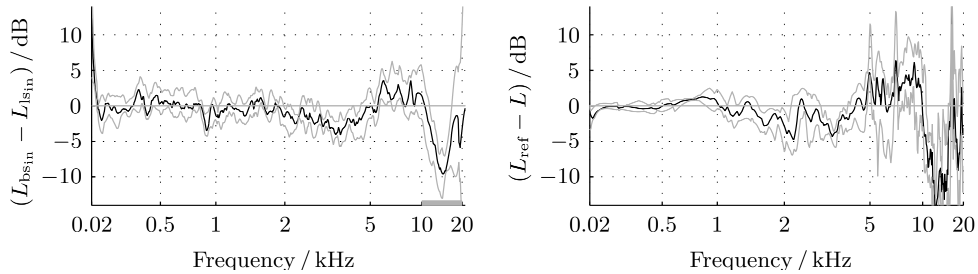

Figure 5: Medians (black) and inter-quartile ranges (gray) of individual results: a) Left panel: Level difference between the input signals to a binaurally synthesized and the corresponding real loudspeaker in front of the subjects in a reverberant laboratory, adjusted to equal loudness by Békésy-Tracking. Dynamic synthesis, non-individual recording, average magnitude equalization; data from Völk and Fastl (2011). b) Right panel: Level difference to the fully individualized situation.

Figure 5 shows the results using a BS system with nonindividual recording and average magnitude equalization. Both graphics are structurally comparable, with similar magnitudes and frequency ranges of the pronounced frequency dependencies. The somewhat larger inter-quartile range at low frequencies in the left panel can be attributed to the accuracy of the tracking procedure, whereas the right panel shows results of physical measurements with higher low-frequency accuracy (Völk, 2011a). However, the physically acquired measurement results shown in the right panel predict the psychoacoustically measured data of the left panel quite well regarding magnitude and frequency, confirming our hypotheses.

The results for BS with individual recording and the corresponding average magnitude equalization are shown by figure 6. Here, the high frequency behavior is again predicted rather well, whereas the tendency towards positive values in the frequency range between 1.5 and 2 kHz appears to be of methodical nature, confirming an hypothesis of Völk and Fastl (2011).

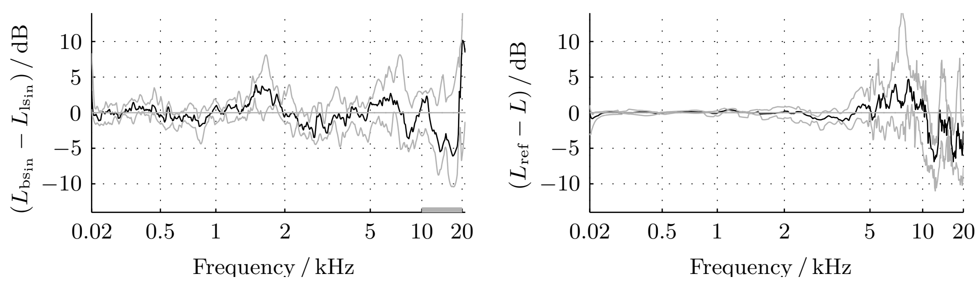

Figure 6: Medians (black) and inter-quartile ranges (gray) of individual results: a) Left panel: Level difference between the input signals to a binaurally synthesized and the corresponding real loudspeaker in front of the subjects in a reverberant laboratory, adjusted to equal loudness by Békésy-Tracking. Static synthesis, individual recording, average magnitude equalization; data from Völk and Fastl (2011). b) Right panel: Level difference to the fully individualized situation.

Figure 7 shows the comparison for synthesis with individual recording and individual magnitude equalization. The prediction appears to be valid for frequencies below about 10 kHz, whereas the high frequency artifact visible in the very high frequency range between 10 and 20 kHz in the left panel is not predicted by the magnitude spectrum on the right. This effect may be attributed to phase effects, but can also be influenced by the fully individualized reference selected here, which represents the best case available in this study, but shows relatively large high-frequency variability (cf. figure 4).

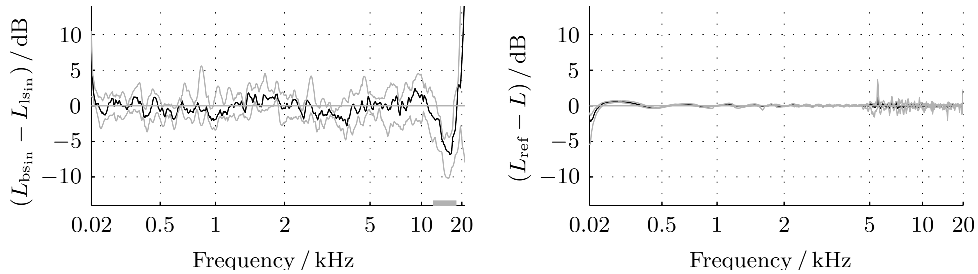

Figure 7: Medians (black) and inter-quartile ranges (gray) of individual results: a) Left panel: Level difference between the input signals to a binaurally synthesized and the corresponding real loudspeaker in front of the subjects in a reverberant laboratory, adjusted to equal loudness by Békésy-Tracking. Static synthesis, individual recording, individual magnitude equalization; data from Völk and Fastl (2011). b) Right panel: Level difference to the fully individualized situation.

CONCLUSIONS AND OUTLOOK

Loudness transfer functions (LTFs), the frequency dependent input level differences between an audio transmission system and the corresponding reference scene at equal loudness for pure tones, have been proposed by Völk and Fastl (2011) as a selective quality measure for audio transmission systems as for example binaural synthesis. It is shown in this paper, that nonindividual spectral ear signal components are visible in the LTFs. This may be interpreted according to Völk (2013) respectively Theile (1980) in that only individual level cues are evaluated in the localization process and consequently not taken into account in forming sound color and loudness, whereas nonindividual level cues influence loudness to the extent they differ from the individual cues. The influence of phase cues on LTFs in BS will be addressed in a future study.

ACKNOWLEDGMENTS

The authors are indebted to Deutsche Forschungsgemeinschaft (DFG) who supported parts of this work by grant FA 140/4 and to Sennheiser electronic GmbH & Co. KG for providing the miniature microphones employed for the measurements.

REFERENECES

Bronkhorst, A. W. (**1995**). “Localization of real and virtual sound sources”, J. Acoust. Soc. Am. **98**, 2542–2553.

Hammershøi, D. and Møller, H. (**1996**). “Sound transmission to and within the human ear canal”, J. Acoust. Soc. Am. **100**, 408–427.

Hudde, H. and Aumann, A. (**2010**). “Zum Begriff des “Trommelfellschalldrucks” – Eine Neubetrachtung unter Verwendung Finiter Elemente (On the term “eardrum sound pressure” – A new discussion using finite elements)”, in *Fortschritte der Akustik, DAGA 2010*, 75–76 (Dt. Gesell. für Akustik e. V., Berlin).

Hudde, H. and Schmidt, S. (**2009**). “Sound fields in generally shaped curved ear canals”, J. Acoust. Soc. Am. **125**, 3146–3157.

Kim, S.-M. and Choi, W. (**2005**). “On the externalization of virtual sound images in headphone reproduction: A wiener filter approach”, J. Acoust. Soc. Am. **117**, 3657–3665.

Kulkarni, A. and Colburn, H. S. (**2000**). “Variability in the characterization of the headphone transfer-function”, J. Acoust. Soc. Am. **107**, 1071–1074.

Mehrgardt, S. and Mellert, V. (**1977**). “Transformation characteristics of the external human ear”, J. Acoust. Soc. Am. **61**, 1567–1576.

Møller, H. (**1992**). "Fundamentals of binaural technology", Applied Acoustics **36**, 171–218.

Møller, H., Hammershøi, D., Jensen, C. B., and Sørensen, M. F. (**1995**a). "Transfer characteristics of headphones measured on human ears", J. Audio Eng. Soc. **43**, 203–217.

Møller, H., Hammershøi, D., Jensen, C. B., and Sørensen, M. F. (**1999**). "Evaluation of artificial heads in listening tests", J. Audio Eng. Soc. **47**, 83–100.

Møller, H., Jensen, C. B., Hammershøi, D., and Sørensen, M. F. (**1996**a). "Selection of a typical human subject for binaural recording", Acustica – Acta Acustica **82**, S 215.

Møller, H., Jensen, C. B., Hammershøi, D., and Sørensen, M. F. (**1996**b). "Using a typical human subject for binaural recording", in *100th AES Convention*, Preprint 4157.

Møller, H., Jensen, C. B., Hammershøi, D., and Sørensen, M. F. (**1997**). "Evaluation of artificial heads in listening tests", in *102nd AES Convention*, Preprint 4404.

Møller, H., Sørensen, M. F., Hammershøi, D., and Jensen, C. B. (**1995**b). "Head-related transfer functions of human subjects", J. Audio Eng. Soc. **43**, 300–321.

Møller, H., Sørensen, M. F., Jensen, C. B., and Hammershøi, D. (**1996**c). "Binaural technique: Do we need individual recordings?", J. Audio Eng. Soc. **44**, 451–469.

Stinson, M. R. (**1985**). "The spatial distribution of sound pressure within scaled replicas of the human ear canal", J. Acoust. Soc. Am. **78**, 1596–1602.

Terhardt, E. (**1985**). "Fourier transformation of time signals: Conceptual revision", Acustica **57**, 242–256.

Theile, G. (**1980**). "Über die Lokalisation im überlagerten Schallfeld (On the localization in the superimposed soundfield)", Ph.D. thesis, Technische Universität Berlin.

Völk, F. (**2010**). "Messtechnische Verifizierung eines datenbasierten binauralen Synthesesystems (Metrological verification of a data based binaural synthesis system)", in *Fortschritte der Akustik, DAGA 2010*, 1049–1050 (Dt. Gesell. für Akustik e. V., Berlin).

Völk, F. (**2011**a). "Inter- and intra-individual variability in blocked auditory canal transfer functions of three circum-aural headphones", in *131st AES Convention*, Convention Paper 8465.

Völk, F. (**2011**b). "System theory of binaural synthesis", in *131st AES Convention*, Convention Paper 8568.

Völk, F. (**2012**a). "Headphone selection for binaural synthesis with blocked auditory canal recording", in *132nd AES Convention*, Convention Paper 8677.

Völk, F. (**2012**b). *Motor- und Aggregate-Akustik IV (Motor and aggregate acoustics IV)*, chapter Audio playback for auditory quality evaluations – Requirements, possibilities, and the impact on applications, 31–47 (Expert, Renningen).

Völk, F. (**2013**). "Interrelations of virtual acoustics and hearing research by the example of binaural synthesis", Ph.D. thesis, Technische Universität München.

Völk, F., Dunstmair, A., Riesenweber, T., and Fastl, H. (**2011**a). "Bedingungen für gleichlaute Schalle aus Kopfhörern und Lautsprechern (Requirements for equally loud sounds from headphones and loudspeakers)", in *Fortschritte der Akustik, DAGA 2011*, 825–826 (Dt. Gesell. für Akustik e. V., Berlin).

Völk, F. and Fastl, H. (**2011**). "Locating the missing 6 dB by loudness calibration of binaural synthesis", in *131st AES Convention*, Convention Paper 8488.

Völk, F., Landsiedel, C., and Fastl, H. (**2011**b). "Auditory adapted exponential transfer function smoothing (AAS)", in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 49–52.

Wenzel, E. M., Arruda, M., Kistler, D. J., and Wightman, F. L. (**1993**). "Localization using nonindividualized head-related transfer functions", J. Acoust. Soc. Am. **94**, 111–123.

Wenzel, E. M., Wightman, F. L., and Kistler, D. J. (**1991**). "Localization with non-individualized virtual acoustic display cues", in *SIGCHI Conference on Human Factors in Computing Systems*, 351–359.