Author Accepted Manuscript (contents identical to the full published version). This paper was presented at the 131st Convention of the Audio Engineering Society 2011, as paper number 8488. The full published version can be found at http://www.aes.org/e-lib/browse.cfm?elib=16014.

Locating the Missing 6 dB by Loudness Calibration of Binaural Synthesis

Florian Völk and Hugo Fastl

AG Technische Akustik, MMK, Technische Universität München, Arcisstraße 21, 80333 München, Germany

Correspondence should be addressed to Florian Völk (florian.voelk@mytum.de)

ABSTRACT

Binaural synthesis is a sound reproduction technology based on the convolution of sound signals with impulse responses defined between a source and a listener's eardrums. The convolution products are typically presented by headphones. For perceptual verification, subjects traditionally remove the headphones to listen to the corresponding real scenario, which is cumbersome and requires a pause between the stimuli. In this contribution, loudness adjustments are presented using a method that allows for direct comparison by defining the reference scene for a listener wearing headphones. Influences of different headphones and equalization procedures are discussed and an explanation for the difference in auditory canal pressure between headphone and loudspeaker reproduction at the same loudness commonly referred to as the missing 6 dB is deduced.

1. INTRODUCTION

In 1989, Wightman and Kistler presented the idea to convolve impulse responses describing the transfer paths from a sound source in the free sound field to the sound pressures at a listener's eardrums (ear signals) with a sound signal and play back the convolution products by headphones (HPs). This way, they intended to synthetically generate the ear signals that would have been elicited by the considered sound source driven with the signal used. The feasibility of this procedure has been proofed by Wightman and Kistler (1989b). It is possible to extended this approach in a straightforward manner to listening situations in reflective environments (cf. e. g. Møller 1992). Adding a head tracking mechanism that controls an adaptive signal processing routine which in turn adjusts the impulse responses to the current listening situation results in a system usually referred to as dynamic data based binaural synthesis (BS), described for example by Foster (1992) and Völk et al. (2007). Different factors are known to influence the resulting auditory events' authenticity, for example the measurement and selection of the impulse responses (cf. Wenzel 1992, Wenzel et al. 1993, Völk et al. 2009) or the amount of reverberation contained in them (cf. Begault 1992, Völk et al. 2008, Völk 2009). BS can be realized with impulse responses recorded either with an artificial or human head. Recording in a human subject's auditory canals can be done with the subject the synthesis is intended for (*individual recording*) or a different subject (*non-individual recording*). The particular recording procedure is known to influence the localization of the so-called virtual sources (cf. e. g. Völk et al. 2008, Völk 2009).

Other elements showing remarkable impact on the synthesis quality achieved are the headphones (cf. Völk 2010), along with their equalization. Considering the latter, Pralong and Carlile (1996) found inter-individual differences in circum-aural headphones' transfer functions measured in ten subjects' auditory canals, and showed their influence on the resulting ear signals as well as the need for equalization. Comparable data were presented by Møller et al. (1995). Pralong and Carlile (1996) reported an increase in the ear signal distortion introduced by non-individual equalization compared to the individually equalized case. Kim and Choi (2005) confirmed in listening experiments the importance of individual equalization for out-of-the-head localization. Summing up, the influence of the components involved in BS on the localization of the virtual sources is well-known. However, there is relatively little known about the influence of the different components on the loudness perception elicited by virtual sources. This issue becomes apparent also when considering the difference in loudness perception between loudspeaker and headphone reproduction at the same sound pressure level in the auditory canals, reported by different authors (e. g. Munson and Wiener 1952, Fastl 1986, Theile 1986). This difference is usually referred to as the *Effect of the Missing 6 dB*.

It is the aim of this paper to compare the loudness elicited by a binaurally synthesized virtual loudspeaker (LS) to the loudness of the corresponding real counterpart by adjustment. The influences of different impulse response measurement and equalization procedures are addressed in detail, resulting in an explanation of the *Effect of the Missing 6 dB*. The paper is structured as follows: Initially, earlier studies regarding the *Effect of the Missing 6 dB* are discussed. Then, aspects of BS are summarized to the extent necessary here. On that basis, the experimental setup and method are introduced, followed by presentation of the results. A discussion and a concluding summary complete the paper.

2. THE EFFECT OF THE MISSING 6 dB

In *Acoustic Measurements*, Leo L. Beranek constitutes supra-aural HPs to require 6 to 10 dB more level at the eardrums to elicit the same loudness perception as a free soundfield (Beranek 1949, pp. 730/731), while this effect's origin being largely unclear. The differences occur for threshold measurements (cf. e. g. Sivian and White 1933), usually referred to as difference between minimum audible field and minimum audible pressure, as well as for supra-threshold loudness adjustments. While the differences at threshold can be attributed largely to methodical shortcomings (see Rudmose 1950, Killion 1978, Rudmose 1982), the effect for levels above threshold persists.

In 1952, Munson and Wiener presented an article titled *In Search of the Missing 6 Db*, reporting the effect especially at low frequencies for diotic HP presentation and binaural listening to a LS in a free sound field. As one difference between the two playback methods and therefore a possible reason for the unexpected deviation, Munson and Wiener mention the different auditory event positions, while not confirming this hypothesis. In 1956, the effect was recognized by Robinson and Dadson while measuring equal loudness contours (cf. Robinson and Dadson 1956), and has further been reported for supra-threshold measurements by Theile (1986).

Rudmose (1982) claims to have found a setup and procedure for the loudness adjustment, resulting in no difference in auditory canal level at equal loudness and considers the *Effect of the Missing 6 dB* to be closed. Possible explanations are according to Rudmose structure-borne sound transmission from the electro-acoustic transducers to the subjects' chair, the LS position, transducer distortions, and the procedure used. Rudmose found further for some subjects a LS close to one ear to require more level for equal loudness than a distant LS, in other words that the LS position can influence the loudness adjustment results.

Fastl et al. (1985) give quantitative differences between the levels of tones in the auditory canals at equal loudness for binaural listening to a LS ($L_{\mathrm{ls_{in}}}$) and diotic HP ($L_{\mathrm{hp_{in}}}$) presentation at levels around 70 dB. Figure 1 shows the quartiles of eight subjects' results according to Fastl et al. (1985, filled circles indicate closed, squares open HPs).

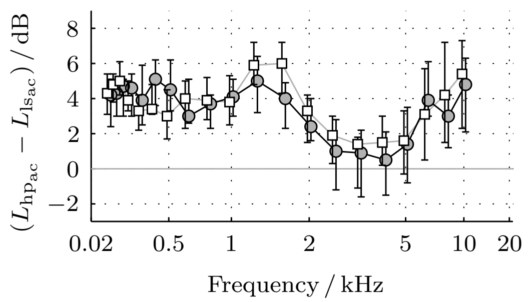

Fig. 1: Quartiles of the level differences in the auditory canal between equally loud tones presented diotically by headphones and a loudspeaker at 3.5 m distance for binaural listening in the anechoic chamber (according to Fastl et al. 1985). Level at the listening position 70 dB SPL. Circles indicate closed, squares open headphones.

Qualitatively well in line with Figure 1 and quantitatively even more pronounced, Keidser et al. (2000) find in their own data as well as in a broad literature review on average 8 dB more level required for HP than for LS playback to elicit the same loudness for sounds in the frequency range around 500 Hz and no difference around 3 kHz. These authors also mention the LS position as a possible reason.

3. SETUP AND PROCEDURE

BS is a modern audio playback procedure, attempting to recreate the sound pressure signals at the eardrums, that would have been present in a (possibly hypothetical) reference situation. To achieve this goal as closely as possible when using HPs for playback, it is required to record the impulse responses involved with miniature microphones at the entrance to the blocked auditory canals and to use appropriate HPs (cf. Møller 1992, Völk 2010).

It is shown in Völk (2010) that a system implemented correctly exhibits on an artificial head a magnitude transfer function to the reference scene independent of frequency (within the accuracy given by the artificial head measurement procedure). This necessarily means equal sound pressure levels for LS and HP playback in the artificial head situation considered. Comparable measurements with a real head are severely limited, if possible at all, since the correct acquirement of the sound pressure distribution across the eardrum resembles a difficult task (cf. e. g. Hudde and Schmidt 2009).

A common way to perceptually assess an electro-acoustical transmission system's properties is the loudness comparison to the reference scene. The frequency dependent level correction necessary for tones to sound as loud as the reference is referred to as the *loudness transfer function* of the system under consideration here. Most prominent loudness transfer functions are the free and diffuse field transfer characteristics of HPs (cf. e. g. Villchur 1969, Fastl and Fleischer 1978, Theile 1986, Møller 1992). Within this paper, loudness transfer functions of different BS implementations are presented.

3.1. Binaural Synthesis

According to the preceding paragraph, the preferable approach to BS is individual miniature microphone recording at the entrances to the blocked auditory canals. Since the recording and the design of HP equalization filters for each subject is time consuming, completely or partially non-individualized procedures are employed often.

Non-individual recording means that the impulse responses are measured on a human subject which is not among the subject group the reproduction is done for. For *individual equalization*, the listener's headphone transfer functions are inverted to form the target function for the equalization filter design. To obtain a filter useful in practical implementations, amplification requirements given by the filter target which can not be provided by the equipment used are reduced by high- and low-pass filtering and regularization (cf. e. g. Kirkeby and Nelson 1999, Norcross et al. 2004). If individual equalization is not appropriate, a target for equalization suitable for different subjects can be acquired by averaging over the magnitude transfer functions of a certain number of subjects. The resulting *average magnitude equalization* filter is then obtained by regularization and combination with a suitable linear phase response here. It is also possible to equalize only the transfer function magnitude individually (*individual magnitude equalization*). The least elaborate method of equalization discussed here is to use the equalization filter designed to account for the magnitude transfer function of a specific subject or an artificial head for others (*non-individual magnitude equalization*).

BS systems that use adaptive signal processing to adjust the impulse responses to the current listening situation are referred to as *dynamic binaural synthesis*, in contrast to *static binaural synthesis*. The dynamic BS used for the loudness adjustments here is restricted to rotational head movements. That means, one set of 360 pairs of impulse responses is used for the synthesis, measured for a rotation of the considered subject. Translational head movements cause no ear signal adaption. This situation resembles an idealized case, not representing a real situation completely. It is shown in section 4 that this situation simulates the ear signals of the real scenario sufficiently for the analysis done here.

3.2. Listening Rooms

In the following, experiments carried out in three different rooms, two laboratories and one anechoic chamber, are presented. Figure 2 shows the corresponding reverberation times, measured at the listening position using the reference scene LS at its reference position as sound source.

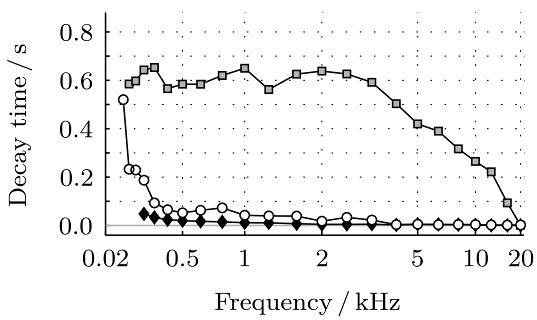

Fig. 2: Third-octave band reverberation times (early decay times) of the rooms where the listening experiments took place: Laboratory 1 (circles), laboratory 2 (squares), anechoic chamber (diamonds).

As to be expected, the anechoic chamber (diamonds) shows almost no reverberation for frequencies above about 250 Hz. Laboratory 1 (circles) is designed to resemble a highly damped living room and therefore provides a low reverberation time too, especially at high frequencies. Laboratory 2 (squares) is a standard small laboratory room with a typical reverberation time of about 600 ms at low and mid frequencies, decreasing towards higher frequencies. Different room acoustical conditions were included to address possibly occurring differences between the loudness transfer functions acquired under reverberant (laboratory 2), damped (laboratory 1), and anechoic room acoustical conditions.

3.3. Setup

In every room, a chair is positioned in front of a LS (Klein + Hummel O 98) at 1.5 m distance. The complete setup is centered around the room midpoint, while care is taken not to position the chair or the LS directly at the midpoint. Initially, in each condition, the impulse responses required for the BS are measured. For the measurements as well as the listening experiments, an optical position control based on the principle of a pinhole camera is used to adjust the chair so that the midpoint of the subject's inter-aural axis lies horizontally and vertically with an accuracy of ± 3 cm on the radiation axis of the LS, for the subject seated in the chair.

For the measurements, the chair is rotated around the midpoint of the inter-aural axis in 6° steps. The intended rotational center and step size is assured using a head tracking system (Polhemus 3 Space FasTrack) with an accuracy of ± 1 cm and $\pm 0.5^\circ$ respectively. In addition, measurements with head nodding or tilting deviations from the horizontal plane of more than $\pm 0.75^\circ$ are repeated until the requirements are met. After the measurements, the horizontal impulse response grid resolution is increased to 1° by cubic spline interpolation over the time shifted impulse responses (cf. Christensen et al. 1999). All experiments are done with this BS resolution.

The measurements are carried out using exponential sine sweeps (ESS, cf. Farina 2000, Müller and Massarani 2001, 5 s duration, frequency range from 10 Hz to 22050 Hz) with the measurement system presented in detail in Völk et al. (2009). The signal processing is carried out at double precision word length with 44.1 kHz sample rate. For all recordings, Sennheiser KE 4-211-2 electret microphones embedded in modified foam ear plugs are inserted in the auditory canals so that the canals are completely blocked and the microphones are positioned some millimeters inside the canal.

The experiments are carried out using two different HP models typically employed for BS (Stax λ pro NEW and Sennheiser HD 800). The respectively used model is indicated along with the results. Since for each model, only one specimen is considered, the results may not be representative for the respective headphone model, but can without loss of generalization be regarded as indication for the necessity to select headphones for binaural playback carefully (cf. Völk 2010). No head fixation is applied and the subjects are allowed to turn their heads. To reduce visual influences, the experiments are carried out in complete darkness.

3.4. Method and Stimuli

The experimental method employed for the loudness adjustments is Békésy-Tracking according to Fastl and Zwicker (2007, cf. also von Békésy 1947, Zwicker and Feldtkeller 1955). In the present case, the subjects listen in turn to a real and the corresponding binaurally synthesized LS without taking off the HPs. It is their task to continuously adjust the level of the BS so that equal loudness is elicited by the synthesis and the real LS. For that purpose, pulsed tones are presented alternately by the BS and the LS. After each pair, the frequency is changed automatically. The level $L_{\mathrm{ls_{in}}}$ at the LS input remains constant while the level $L_{\mathrm{bs_{in}}}$ of the BS is either increased or decreased after each pair. The subject can change the direction of the level variation using a hand switch and is asked to cause a change in direction every time the loudness of the two tones in one pair differs. This procedure results in a frequency dependent zigzag-pattern, alternating around the line of equal loudness.

For the study on hand, tone impulses with 0.4 s pulse duration, 5 ms Gaussian gating, and 0.1 s pause between the two tones to be compared are used. Two successive pairs are separated by 0.4 s pause. The second tone, presented by the BS system, is adjusted to a reference tone of the same frequency, presented by the LS (calibrated with broad band noise to a level of about 58 dB SPL). The BS level is varied with 1.5 dB step size, starting 10 dB above the level eliciting approximately the loudness of the LS.

Each experimental run is divided in two parts, one with increasing frequency (increase equally spaced on the auditory adequate critical-band rate scale, cf. Fastl and Zwicker 2007), starting from 10 Bark (about 1.3 kHz) upwards to 24.6 Bark (about 20 kHz), and one with decreasing frequency, starting from 12 Bark (about 1.7 kHz) downwards to 0.2 Bark (about 20 Hz) at a step size of 0.05 Bark. To reduce methodical artifacts due to the beginning of the adjustment procedure, 20 steps of the results in the overlapping region, that is between 10 and 11 Bark for the increasing and 12 and 11 Bark for the decreasing branch, are dropped. The individual result for one subject is then computed as interpolation over the mean values between every two neighboring turning points. Using this procedure, a frequency dependent mean deviation between two runs of about ±2 dB is achieved (measured for three experienced subjects, maximum deviation ±4 dB).

4. RESULTS AND DISCUSSION

In this section, the results of the loudness adjustments (the loudness transfer functions) of a binaurally synthesized to the corresponding real LS are discussed for different recording conditions, HPs, and equalization methods. The order of presentation is selected so that the discussion of each situation is based on the preceding data. At least eight subjects participated in all experiments. The results for each condition are computed as inter-individual medians and inter-quartile ranges of the individual loudness transfer functions in the respective condition.

4.1. Non-Individual Recording, Average Magnitude Equalization

A typical use case of BS is non-individual recording with average magnitude equalization. There are three possible synthesis errors in that situation: non-individual measurement, average HP equalization, and non-appropriate HPs.

4.1.1. Stax λ pro NEW Headphones

Figure 3 shows the loudness transfer function (the level difference between the binaurally synthesized and the real LS inputs $\Delta L = L_{\mathrm{bs_{in}}} - L_{\mathrm{ls_{in}}}$) obtained in the reflective laboratory room 1 with non-individual recording and average magnitude equalization using Stax λ pro NEW HPs, which are often used for BS applications (cf. e. g. Møller et al. 1995, Spikofski and Fruhmann 2001).

The results suggest that the loudness perception elicited by the LS is well preserved by the BS for frequencies up to about 6 kHz. In the frequency range between 6 and 10 kHz, the BS system must be driven on average with about 6 dB more level than the LS to elicit equal loudness (median deviations above the methodical accuracy are highlighted by gray bars on the x-axis). At frequencies between 12 and 15 kHz, the median of the level difference decreases, while at the upper limit of the audible frequency range, again higher levels are necessary for the BS to elicit the same loudness as the LS.

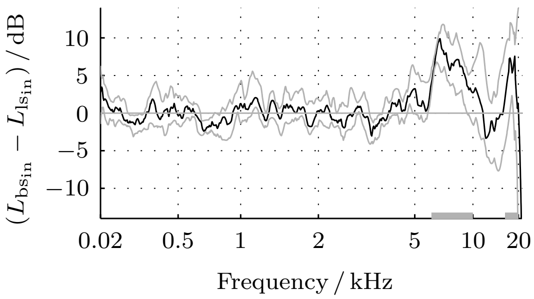

Fig. 3: Median and inter-quartile range of the individual level differences between the input signals of a binaurally synthesized and the corresponding real loudspeaker in front of the subjects in a reflective laboratory (room 1), adjusted to equal loudness by Békésy-Tracking. Headphone specimen Stax λ pro NEW, Dynamic synthesis, non-individual recording, average magnitude equalization.

Since the loudness of a fixed frequency sinusoidal with 400 ms duration depends solely on its level (cf. Fastl and Zwicker 2007), it should be possible to achieve equal loudness perception by adjusting the BS input level according to Figure 3. Figure 4 shows the loudness adjustment results for the BS setup described above with additional initial level correction according to the median shown in Figure 3. The resulting level difference basically resembles the situation without input level correction, while the global characteristics are somewhat more flattered, which might be due to the steeper level changes caused by the initial level correction.

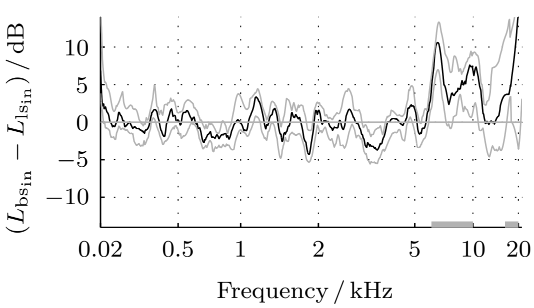

Fig. 4: Median and inter-quartile range of the individual level differences between the input signals of a binaurally synthesized and the corresponding real loudspeaker in front of the subjects in room 1, adjusted to equal loudness by Békésy-Tracking. Headphone specimen Stax λ pro NEW, Dynamic synthesis, non-individual recording, avg. magnitude equalization.

The fact that the deviation can not be equalized by level correction leads to the assumption that resonances in the HP/ear system are the cause of the deviation. Considering the HP selection criterion presented in Völk (2010), which is derived exemplary for both HP specimen also discussed in this paper, the specimen of Sennheiser HD 800 HPs studied in Völk (2010) is more likely to produce frequency independent loudness transfer functions when used in BS than the considered Stax λ pro NEW specimen. It is shown in Völk (2010) that the artificial head magnitude transfer functions of a BS system with regard to the reference situation are frequency independent when using the Sennheiser HD 800 specimen while depend on frequency in the range above about 6 kHz when using the Stax λ pro NEW specimen. Therefore, the Sennheiser HD 800 specimen considered is expected to produce loudness transfer functions that depend less on frequency than those acquired using the Stax λ pro NEW specimen. This hypothesis is validated in the following section.

4.1.2. Sennheiser HD 800 Headphones

Figure 5 shows the loudness transfer function of non-individualized BS implemented using Sennheiser HD 800 HPs in laboratory 2 (cf. Figure 2).

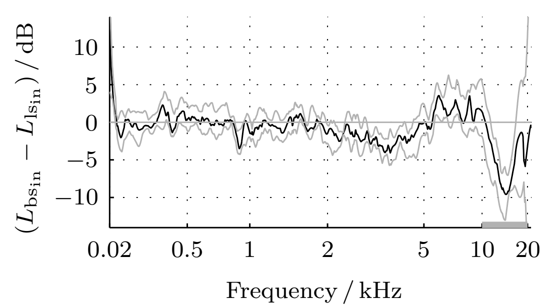

Fig. 5: Median and inter-quartile range of the individual level differences between the input signals of a binaurally synthesized and the corresponding real loudspeaker in front of the subjects in a reflective laboratory (room 2), adjusted to equal loudness by Békésy-Tracking. Headphone specimen Sennheiser HD 800, dynamic synthesis, non-ind. recording, avg. magnitude equalization.

The results shown in Figure 5 indicate that non-individualized BS with Sennheiser HD 800 HPs is capable of approximately reproducing the loudness elicited by the real LS for frequencies below some 10 kHz, apart from a reduced loudness of the BS for very low frequencies (below about 100 Hz). The highly increased level necessary for the BS using the Stax λ pro NEW specimen between 6 and 10 kHz is reduced, whereas between 10 and 20 kHz the BS input level is adjusted lower than expected when using the Sennheiser HD 800 HPs.

Since it can not be excluded that the acoustical characteristics of the reproduction room influence the loudness adjustment results, the experiment is repeated with the otherwise unmodified setup in an anechoic chamber (cf. Figure 6).

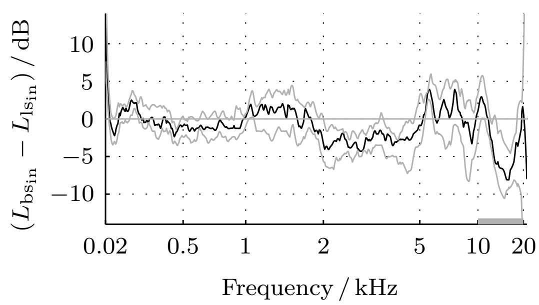

Fig. 6: Median and inter-quartile range of the individual level differences between the input signals of a binaurally synthesized and the corresponding real loudspeaker in front of the subjects in an anechoic chamber, adjusted to equal loudness by Békésy-Tracking. Headphone specimen Sennheiser HD 800, Dynamic synthesis, non-individual recording, average magnitude equalization.

If Figures 5 and 6 are compared, the largest changes occur in the frequency range between 1 and 2 kHz that is in the starting region of the tracking procedure. These changes are attributed to procedural reasons. All subjects reported the adjustment procedure in the anechoic chamber to be more demanding and the listening situation to be less familiar than in the reflective laboratory 2. As a consequence, the subjects need more time and by that a wider frequency range to adjust the starting level to the level at equal loudness under anechoic conditions. The resulting artifact is visible as enlarged inter-quartile range and higher median level in the frequency range between 1 and 2 kHz of the inter-individually averaged data especially for the first experiment conducted in the anechoic chamber (cf. Figure 6), but to a reduced degree also in the second experiment carried out under anechoic conditions (Figure 7). However, the deviations for frequencies between 10 and 20 kHz are comparable in the anechoic chamber (Figure 6) and in laboratory 2 (Figure 5). Consequently, the room acoustical environments show only methodical influence on the loudness transfer functions.

Two possible reasons for the frequency dependence remain: average magnitude equalization and non-individual recording. Therefore, loudness transfer functions for other equalization and synthesis methods are studied systematically in the following, starting with non-individual synthesis and non-individual (section 4.2) as well as individual magnitude equalization (4.3). After showing that static BS allows to study loudness transfer (section 4.4), individual synthesis is evaluated with average (4.5), individual magnitude (4.6), as well as individual magnitude and phase equalization (4.7).

4.2. Non-Individual Synthesis, Non-Individual Magnitude Equalization

To address differences between average and non-individual magnitude equalization, the experiment is repeated in the anechoic chamber with equalization based on the magnitude of the HP transfer functions of the subject used for recording (cf. Figure 7).

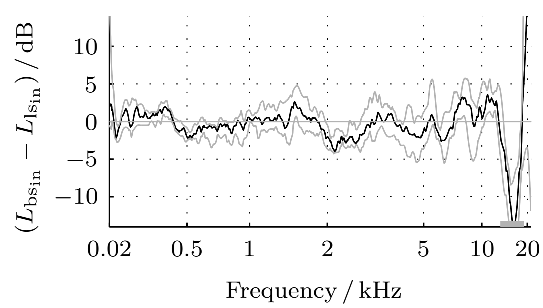

Fig. 7: Median and inter-quartile range of the individual level differences between the input signals of a binaurally synthesized and the corresponding real loudspeaker in front of the subjects in an anechoic chamber, adjusted to equal loudness by Békésy-Tracking. Headphone specimen Sennheiser HD 800, Dynamic synthesis, non-ind. recording, non-ind. magnitude equalization.

According to Figure 7, with non-individual BS and the corresponding non-individual magnitude equalization, the high-frequency deviation becomes narrower than with average magnitude equalization but remains. This means that in combination with non-individual synthesis, neither average nor non-individual magnitude equalization result in frequency independent loudness transfer functions. Since the adjustments are less demanding under reflective conditions, all experiments presented in the following are acquired in laboratory 2.

4.3. Non-Individual Synthesis, Individual Magnitude vs. Average Magnitude Equalization

In Figure 8, a comparison of the medians of the loudness transfer functions in the reflective laboratory 2 using average magnitude (light) and individual magnitude equalization (dark) is given.

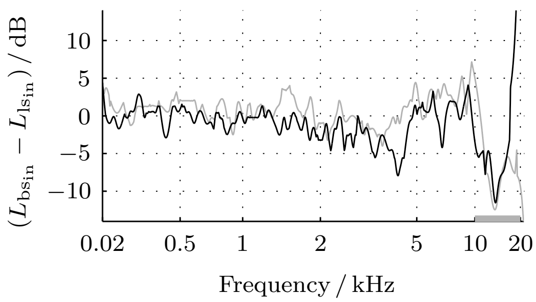

Fig. 8: Median of the individual level differences between the input signals of a binaurally synthesized and the corresponding real loudspeaker in front of the subjects in a reflective laboratory (room 2), adjusted to equal loudness by Békésy-Tracking. Headphone specimen Sennheiser HD 800, Dynamic synthesis, non-individual recording, individual (dark) and average magnitude (light) equalization.

Non-individual recording is not capable of producing frequency independent loudness transfer functions, even with individual equalization. For that reason, individual recording is addressed in sections 4.5 to 4.7. Due to the effort when recording a complete set of individual transfer functions, it would be preferable to use static instead of dynamic BS. This way, the effort could be reduced to one instead of 60 recordings per ear in the configuration regarded here. To justify this procedure, section 4.4 gives a comparison between static and dynamic non-individual BS with average magnitude equalization.

4.4. Dynamic vs. Static Non-Individual Synthesis, Average Magnitude Equalization

Figure 9 gives a comparison of the medians of the loudness transfer functions with static and dynamic BS of the same situation (reflective laboratory 2, non-individual recording and average magnitude equalization).

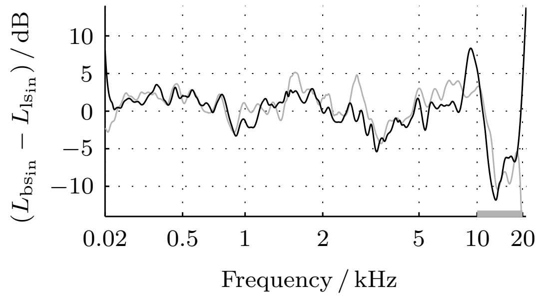

Fig. 9: Median of the individual level differences between the input signals of a binaurally synthesized and the corresponding real loudspeaker in front of the subjects in a reflective laboratory (room 2), adjusted to equal loudness by Békésy-Tracking. Headphone specimen Sennheiser HD 800, Static (dark) and dynamic (light) synthesis, non-individual recording, average magnitude equalization.

Based on the comparison shown in Figure 9, static BS is assumed to resemble the dynamic situation with an average deviation smaller than the average accuracy of the method employed here for measurement of loudness transfer functions (±2 dB, cf. section 3.4). All results presented further in this paper are acquired with static individual BS.

4.5. Individual Synthesis, Average Magnitude Equalization

Figure 10 depicts the result for static individual recording and average magnitude equalization in laboratory 2 with Sennheiser HD 800 headphones.

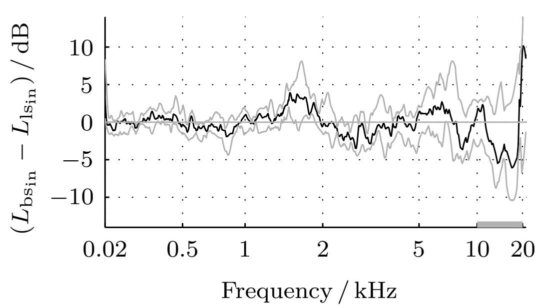

Fig. 10: Median and inter-quartile range of the individual level differences between the input signals of a binaurally synthesized and the corresponding real loudspeaker in front of the subjects in the reflective laboratory 2, adjusted to equal loudness by Békésy-Tracking. Headphone specimen Sennheiser HD 800, Static synthesis, individual recording, average magnitude equalization.

The effect that less level is needed for equal loudness with headphone versus loudspeaker playback between 10 and 20 kHz is visible in the results also for individual binaural synthesis with average equalization. However, the extent of the effect is reduced compared to the non-individual situations. Further reduction is to be expected with individual equalization, as addressed in the next sections.

4.6. Individual Synthesis, Individual Magnitude Equalization

Figure 11 shows the loudness transfer function for static individual BS with Sennheiser HD 800 HPs and individual magnitude equalization in the reflective laboratory room 2.

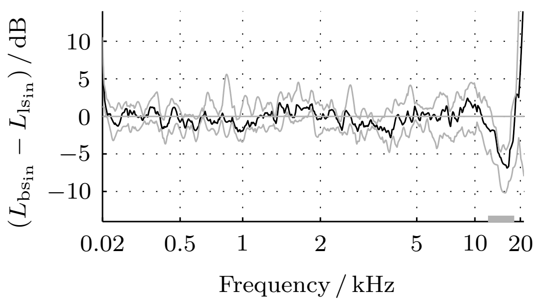

Fig. 11: Median and inter-quartile range of the individual level differences between the input signals of a binaurally synthesized and the corresponding real loudspeaker in front of the subjects in the reflective laboratory 2, adjusted to equal loudness by Békésy-Tracking. Headphone specimen Sennheiser HD 800, Static synthesis, individual recording, individual magnitude equalization.

The overall developing below 12 kHz can be considered frequency independent within the accuracy given by the adjustment method, while the deviation between 12 and 18 kHz is visible also for individual synthesis with individual magnitude equalization. Therefore, individual magnitude and phase equalization is considered in the following section.

4.7. Individual Synthesis, Individual Equalization

In Figure 12, the inter-individual statistics of the loudness transfer functions acquired in the reflective laboratory room 2 with individual BS combined with individual magnitude and phase equalization using Sennheiser HD 800 HPs are depicted.

The data set shown in Figure 12 is considered frequency independent within the accuracy given by the adjustment procedure employed. This can be understood in that individually equalized individual BS is capable of recreating the reference situation loudness over the whole audible frequency range with the accuracy of the measurement procedure.

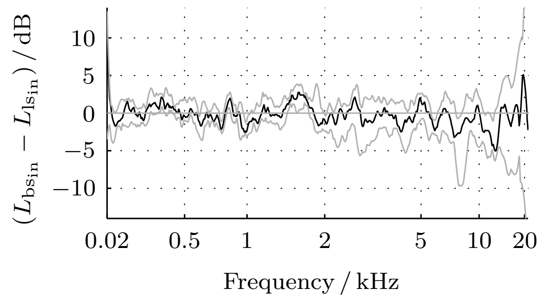

Fig. 12: Median and inter-quartile range of the individual level differences between the input signals of a binaurally synthesized and the corresponding real loudspeaker in front of the subjects in the reflective laboratory 2, adjusted to equal loudness by Békésy-Tracking. Headphone specimen Sennheiser HD 800, Static synthesis, individual recording, individual magnitude and phase equalization.

5. LOCATING THE MISSING 6 dB

Fastl et al. (1985) state that *"typically the sound signals at the eardrums are regarded as the most essential acoustical input parameters leading to auditory sensations in subjects"*, but that *"on the contrary* [...] *tones from loudspeaker versus headphones can be perceived with different loudness despite equal sound level in the auditory canal."* However, the situations contrasted here are indeed not contradictory, since the *sound signals at the eardrums* resemble time dependent variables, which are not described thoroughly by the root-mean-square-value based sound level. Especially different inter-aural phase relations are not detected by the sound level. Obviously, it is accurate to state that *tones at equal level in the auditory canal* can be *perceived with different loudness*. The claim of identical *time functions* of the ear signals in headphone and loudspeaker reproduction on the contrary actually leads to equal loudness perception at equal level in the auditory canal. This is proven valid by the frequency independent loudness transfer function shown for individually equalized individual binaural synthesis in Figure 12, since binaural synthesis aims at recreating the *time signals* occurring in the reference scene.

Comparison of the loudness transfer functions for individual binaural synthesis with individual magnitude equalization (Figure 11) and with individual magnitude and phase equalization (Figure 12) reveals the influence of the (inter-aural) phase relations even for pure tones. These facts can be summarized as follows: Comparing the situations with the same (root-mean-square) sound pressure levels caused by headphone and loudspeaker reproduction does not assure the same ear signals (time dependent sound pressure signals at the eardrums) in both situations. It is well known that the actual sound pressure time function (not only its envelope or average) can influence auditory perceptions (cf. e. g. masking-period-patterns as described by Zwicker 1976, Fastl and Zwicker 2007) or binaural masking level differences (cf. Zwicker and Henning 1991). It seems not to be sufficient to assure the same sound level in the auditory canals to elicit the same loudness in any case (as can be deduced for example from Fastl 1986).

Considering the results of the paper on hand, it can be concluded that the accuracy of the ear signals that can be achieved with individually equalized static individual binaural synthesis is sufficient to elicit equal loudness perception with headphone and loudspeaker presentation without using elaborate methods as described e. g. by Rudmose (1982).

Therefore, based on the experiments presented in this paper, the case of the *Effect of the Missing 6 dB* can be regarded as closed. The explanation for the difference in aclevel for loudspeaker and headphone reproduction at the same sound pressure level is the fact that the same sound pressure time function has to be present at least to the degree achievable with static individual binaural synthesis. The data show in addition that the re-creation of the reference scene ear signals can be assumed as the ultimate goal of binaural synthesis.

The explanation given here was formulated in a more descriptive manner by Theile (1980) as the so called *association principle*. This principle interpreted in view of the problem considered here states that the auditory impression of the loudspeaker includes the perceived source position (the so called auditory event position), which is different in the (traditional) headphone listening situation, since in that case the auditory events are located in the head (lateralization, cf. e. g. Jeffress and Taylor 1961, Plenge 1974). It can be concluded from the association principle that the same loudness perception (and the same overall auditory impression) can only be elicited if the auditory event position is comparable in both cases. The data presented in this paper can be regarded as psychophysical proof of this psychologically motivated hypothesis.

6. CONCLUSIONS

In this study, the loudness elicited by a binaurally synthesized virtual loudspeaker box is compared to the loudness of the corresponding real counterpart. The results indicate that the loudness perception of the loudspeaker box can be approximated over the whole audible frequency range by individual binaural synthesis with individual magnitude and phase equalization using headphones that fulfill the headphone selection criterion proposed in Völk (2010).

Less elaborate binaural synthesis procedures are also capable of eliciting the same loudness as the loudspeaker box for frequencies below about 6 kHz, while creating different perceptions for higher frequencies. The reasons for these deviations can be attributed to non-individual measurement (cf. Figures 5 and 10) and equalization (cf. Figures 10 and 12) as well as inappropriate headphones (cf. Figures 3 and 5 as well as Völk 2010).

The negligible difference between static and dynamic binaural synthesis shown in Figure 9 reveals that *dynamic variation* of the ear signals due to (small) head movements shows no considerable influence on loudness perception in binaural synthesis.

Further, an explanation of the *Effect of the Missing 6 dB* is derived from the experimental results: the same sound-pressure time-functions in the auditory canals ensure the same loudness in loudspeaker and headphone reproduction.

An additional result is the proof that instrumental free-field equalization (cf. e. g. Zwicker and Maiwald 1963, Villchur 1969) is possible by adequately set up static binaural synthesis with impulse responses measured under anechoic conditions, without perceptual loudness comparisons. In case the measurements are carried out in a reverberant chamber, instrumental diffuse-field equalization (cf. e. g. Theile 1986) is possible, too.

7. ACKNOWLEDGMENTS

The authors are indebted to Prof. Dr.-Ing. H. Fleischer for the possibility to carry out the measurements in the anechoic chamber of Universität der Bundeswehr, München Neubiberg.

Further, the help of Sennheiser electronic GmbH & Co. KG who provided the miniature microphones for the measurements is gratefully acknowledged.

Parts of this work were supported by grant FA 140/4 of the Deutsche Forschungsgemeinschaft (DFG).

8. REFERENCES

Begault D. R.: Perceptual Effects of Synthetic Reverberation on Three-Dimensional Audio Systems. J. Audio Eng. Soc. **40**, 895–904 (1992)

Beranek L. L.: *Acoustic Measurements*. (John Wiley & Sons, Inc., New York, 1949)

Christensen F., H. Møller, P. Minnaar, J. Plogsties, S. K. Olesen: Interpolating between Head-Related Transfer Functions Measured with Low Directional Resolution. In *107th AES Convention* (1999) (Preprint 5047)

Farina A.: Simultaneous Measurement of Impulse Response and Distortion with a Swept-Sine Technique. In *108th AES Convention* (2000) (Preprint 5093)

Fastl H.: Gibt es *den* Frequenzgang von Kopfhörern? (Does *the* frequency response of headphones exist?). In *NTG-Fachberichte 91, Hörrundfunk 7*, 274–281 (VDE-Verlag, Berlin, Offenbach, 1986)

Fastl H., H. Fleischer: Freifeldübertragungsmaße verschiedener elektrodynamischer und elektrostatischer Kopfhörer (Free-field transfer functions of different electrodynamic and electrostatic headphones). Acustica **39**, 182–187 (1978)

Fastl H., E. Zwicker: *Psychoacoustics – Facts and Models*. 3rd Edition (Springer, Berlin, Heidelberg, 2007)

Fastl H., W. Schmid, G. Theile, E. Zwicker: Schallpegel im Gehörgang für gleichlaute Schalle aus Kopfhörern oder Lautsprechern (Sound level in the ear canal for equally loud sounds from headphones versus loudspeakers). In *Fortschritte der Akustik, DAGA '85*, 471–474 (DPG, Bad Honnef, 1985)

Foster S.: Real-time implementation of spatial auditory displays. J. Acoust. Soc. Am. **92**, 2332 (1992)

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

Jeffress L. A., R. W. Taylor: Lateralization vs Localization. J. Acoust. Soc. Am. **33**, 482–483 (1961)

Keidser G., R. Katsch, H. Dillon, F. Grant: Relative loudness perception of low and high frequency sounds in the open and occluded ear. J. Acoust. Soc. Am. **107**, 3351–3357 (2000)

Killion M. C.: Revised estimate of minimum audible pressure: Where is the "missing 6 dB"? J. Acoust. Soc. Am. **63**, 1501–1508 (1978)

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

Kirkeby O., P. A. Nelson: Digital Filter Design for Inversion Problems in Sound Reproduction. J. Audio Eng. Soc. **47**, 583–595 (1999)

Møller H.: Fundamentals of Binaural Technology. Applied Acoustics **36**, 171–218 (1992)

Møller H., D. Hammershøi, C. B. Jensen, M. F. Sørensen: Transfer Characteristics of Headphones Measured on Human Ears. J. Audio Eng. Soc. **43**, 203–217 (1995)

Müller S., P. Massarani: Transfer-Function Measurement with Sweeps. J. Audio Eng. Soc. **49**, 443–471 (2001)

Munson W. A., F. M. Wiener: In Search of the Missing 6 Db. J. Acoust. Soc. Am. **24**, 498–501 (1952)

Norcross S. G., G. A. Soulodre, M. C. Lavoie: Subjective Investigations of Inverse Filtering. J. Audio Eng. Soc. **52**, 1003–1027 (2004)

Plenge G.: On the differences between localization and lateralization. J. Acoust. Soc. Am. **56**, 944–951 (1974)

Pralong D., S. Carlile: The role of individualized headphone calibration for the generation of high fidelity virtual auditory space. J. Acoust. Soc. Am. **100**, 3785–3793 (1996)

Robinson D. W., R. S. Dadson: A re-determination of the equal-loudness relations for pure tones. Brit. J. of Appl. Physics **7**, 166–181 (1956)

Rudmose W.: Free-Field Thresholds vs. Pressure Thresholds at Low Frequencies. J. Acoust. Soc. Am. **22**, 674 (1950)

Rudmose W.: The case of the missing 6 dB. J. Acoust. Soc. Am. **71**, 650–659 (1982)

Sivian L. J., S. D. White: On Minimum Audible Sound Fields. J. Acoust. Soc. Am. **4**, 288–321 (1933)

Spikofski G., M. Fruhmann: Optimisation of Binaural Room Scanning (BRS): Considering inter-individual HRTF-characteristics. In *19th Int. AES Conference* (2001) (Paper Number 1907)

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

Theile G.: On the Standardization of the Frequency Response of High-Quality Studio Headphones. J. Audio Eng. Soc. **34**, 956–969 (1986)

Villchur E.: Free-Field Calibration of Earphones. J. Acoust. Soc. Am. **46**, 1527–1534 (1969)

Völk F.: Externalization in data-based Binaural Synthesis: Effects of Impulse Response Length. In *Int. Conference on Acoustics NAG/DAGA 2009*, 1075–1078 (Dt. Gesell. für Akustik e. V., Berlin, 2009)

Völk F.: 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, 2010)

Völk F., F. Heinemann, H. Fastl: Externalization in binaural synthesis: effects of recording environment and measurement procedure. In *Acoustics '08*, 6419–6424 (2008)

Völk F., S. Kerber, H. Fastl, S. Reifinger: Design und Realisierung von virtueller Akustik für ein Augmented-Reality-Labor (Design and implementation of virtual acoustics for an augmented reality lab). In *Fortschritte der Akustik, DAGA 2007*, 673–674 (Dt. Gesell. für Akustik e. V., Berlin, 2007)

Völk F., M. Straubinger, L. Roalter, H. Fastl: Measurement of Head Related Impulse Responses for Psychoacoustic Research. In *Int. Conference on Acoustics NAG/DAGA 2009*, 164–167 (Dt. Gesell. für Akustik e. V., Berlin, 2009)

von Békésy G.: A new audiometer. Acta Otolaryngol. **35**, 411–422 (1947)

Wenzel E. M.: Issues in the development of virtual acoustic environments. J. Acoust. Soc. Am. **92**, 2332 (1992)

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

Wightman F. L., D. J. Kistler: Headphone simulation of free-field listening. I: Stimulus synthesis. J. Acoust. Soc. Am. **85**, 858–867 (1989a)

Wightman F. L., D. J. Kistler: Headphone simulation of free-field listening. II: Psychophysical validation. J. Acoust. Soc. Am. **85**, 868–878 (1989b)

Zwicker E.: Masking period patterns of harmonic complex tones. J. Acoust. Soc. Am. **60**, 429–439 (1976)

Zwicker E., R. Feldtkeller: Über die Lautstärke von gleichförmigen Geräuschen (On the loudness of uniform sounds). Acustica **5**, 303–316 (1955)

Zwicker E., G. B. Henning: On the effect of interaural phase differences on loudness. Hearing Research **53**, 141–152 (1991)

Zwicker E., D. Maiwald: Über das Freifeldübertragungsmaß des Kopfhörers DT 48 (On the free-field response of the earphone DT 48). Acustica **13**, 181–182 (1963)