ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/310600534

Measurement and Perceptual Evaluation of a Spherical Near-Field HRTF Set

Conference Paper · November 2016

CITATIONS
25

READS
1,783

3 authors:

Johannes M. Arend
Technische Universität Berlin
83 PUBLICATIONS **599** CITATIONS

SEE PROFILE

Annika Neidhardt
Technische Universität Ilmenau
44 PUBLICATIONS **234** CITATIONS

SEE PROFILE

Christoph Pörschmann
Technische Hochschule Köln
122 PUBLICATIONS **973** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project Binaural hearing in the real and virtual world to improve school children's hearing experience (VIWER-S) View project

Project Human Voice Directivity View project

All content following this page was uploaded by Johannes M. Arend on 26 February 2017.

The user has requested enhancement of the downloaded file.

Measurement and Perceptual Evaluation of a Spherical Near-Field HRTF Set

(Messung und perzeptive Evaluierung eines sphärischen Satzes von Nahfeld-HRTFs)

Johannes M. Arend[1,2], Annika Neidhardt[3], Christoph Pörschmann[1]
[1] *TH Köln, Institute of Communications Engineering, Cologne, Germany*
[2] *TU Berlin, Audio Communication Group, Berlin, Germany*
[3] *TU Ilmenau, Electronic Media Technology, Ilmenau, Germany*

Abstract

The perceptual refinement of dynamic binaural synthesis has been subject to research for the past years. The basic principle relies on head-related transfer functions (HRTFs), which describe the directional filtering caused by the head, pinna, and torso. However, most systems are based on far-field HRTFs and therefore ignore the acoustical specifics of near-field sound sources. One reason might be that full spherical near-field HRTF sets are rarely available. In this paper, we present an HRTF set of a Neumann KU100 dummy head. The set is freely available for download and contains post-processed impulse responses, captured on a circular and full spherical grid at sound source distances between 0.25 m and 1.50 m. In a subsequent listening experiment using dynamic binaural synthesis, we investigated if the captured binaural cues affect estimated distance of a virtual sound source. The set is useful for various spatial audio applications where nearby virtual sound sources are required, such as auditory displays.

1. Introduction

These days, dynamic binaural synthesis can be regarded as a state-of-the-art approach for headphone-based spatial audio reproduction. The basic principle relies on head-related transfer functions (HRTFs), which describe the directional filtering of the incoming sound caused by the head, pinna, and torso. At this time, a variety of HRTF datasets are available, such as individual HRTF measurements (CIPIC database [1] for example), the established KEMAR dummy head HRTFs [2], or high spatial resolution data of a Neumann KU100 dummy head [3]. The SOFA repository [4] provides an extensive collection of diverse HRTF datasets unified in one data format. In general, the sets are based on anechoic measurements or, in some cases, on simulations. However, most datasets currently available are far-field HRTFs, which means that the sound source used for measurements or simulations was placed at a distance of at least 1 m. Thus, the acoustical specifics of nearby sound sources in the so-called proximal region [5] (the region within 1 m of the listener's head) are simply ignored, even though these features are well known. Stewart [6], Hartley et al. [7], and Brungart et al. [5] for instance theoretically examined the influence of increased head shadowing for nearby sound sources. The studies revealed substantial changes in HRTFs for proximal-region sources. Furthermore, Brungart et al. [5] conducted detailed physical analyses of near-field HRTF data, based on measurements with a KEMAR dummy head. Here, the authors showed a significant increase of interaural level differences (ILDs) as well as an increasing low-pass filtering character of the HRTFs as the sound source approaches the head. Moreover, they outlined the parallax effect for nearby sound sources that especially gains importance when head movements are involved, as is the case with dynamic binaural synthesis. In two subsequent publications, Brungart et al. [8] [9] investigated auditory localization of nearby sound sources. Concerning auditory distance perception in anechoic environments, they conducted a study where subjects had to estimate distance of various level-normalized stimuli, thus loudness-based distance cues were missing. Their results suggested that the specific binaural features found in the HRTFs for nearby sound sources are an important distance cue in the proximal region. As opposed to this, Shinn-Cunningham et al. [10] [11] found in a similar experiment that binaural cues were irrelevant for proximal-region distance perception in anechoic environments. The contrary results show that further investigations in this topic are needed.

Overall, it becomes apparent that the clearly different features of near-field HRTFs should be considered for auralization purposes. Near-field HRTFs for virtual nearby sound sources might improve the plausibility of the virtual auditory scene. Furthermore, proximal-region effects as well as motion-dependent parallax of virtual nearby sound sources could be implemented satisfactorily. Besides, a set of high resolution near-field HRTFs that is publicly available could be used for further experiments regarding auditory localization of nearby sound sources. So far, there are only a few datasets available, whereby some of them can be freely accessed on the Internet [12] [13] [14] and others hardly can be found [15] [16] [17]. However, none of these datasets provide HRTFs with a high SNR over the full audible bandwidth, measured on a full spherical grid with high angular resolution. For use in virtual acoustics, a high-resolution full spherical dataset has several advantages. First of all, it provides a high number of discrete measurement points according to the used spatial sampling grid. Moreover, the dataset can be transformed to the spherical harmonic domain. This allows for spherical harmonic interpolation, which is valid on the entire audible spectrum, given that the measurement resolution is high enough (see [18, Chapter 3.12.4]). As a result, any arbitrary near-field HRTF can be obtained for the respective measured sound source distance. Thus, measuring full spherical datasets at several positions in the proximal region covers a wide range of possible near-field HRTFs.

Eight kHz, LLC
Exhibit 2011
IPR2023-01024
Page 2

ISBN 987-3-9812830-7-5

In this paper, we present such a full spherical HRTF database of a Neumann KU100 dummy head, measured with high angular resolution at sound source distances between 0.50 m and 1.50 m. To our knowledge, there is no other full spherical near-field HRTF dataset (of a KU100 dummy head) available so far. Additionally, we captured HRTFs on a circular grid at distances between 0.25 m and 1.50 m, also presented here. The final set is considered to be useful for various auralization applications, like auditory displays or architectural acoustics. Therefore, the focus was on precise positioning, high SNR and full audible bandwidth. Based on the new HRTF set, we conducted several listening experiments. In one experiment, which is presented in this paper, we investigated if the HRTFs can be applied to code distance and if appropriate distance estimation is still possible when natural level differences between the stimuli are missing. The paper is structured as follows. Section 2 describes the HRTF measurements including the measurement setup, the applied post-processing and a technical evaluation of the final HRTF dataset. Section 3 provides the perceptual evaluation of the measured near-field HRTFs. It outlines the used test design as well as preliminary results. Finally, section 4 concludes the paper with a short summary of the measurements and the main findings.

2. HRTF Measurements

2.1. Setup

The HRTF measurements were performed in the anechoic chamber of the acoustics laboratory at TH Köln. The chamber has dimensions of 4.5 m × 11.7 m × 2.30 m (W×D×H) and a low cut-off frequency of about 200 Hz. The sound source was a Geithain RL906 loudspeaker, which has a two-way coaxial design and a flat on-axis magnitude response from 50 Hz to 20 kHz (±3 dB). Thus, the loudspeaker approaches the ideal of an acoustic point source and allows measuring HRTFs in almost the full audible frequency range. The VariSphear measurement system [19] was used for precise positioning of the Neumann KU100 dummy head at the spatial sampling positions and for capturing head-related impulse responses (HRIRs), which is the time-domain equivalent of HRTFs. The impulse responses were measured according to two different spatial sampling grids with two different VariSphear setups: a circular grid where the dummy head was fully rotated in the horizontal plane in steps of 1°, and a Lebedev full spherical grid with 2702 points. The latter is well suited for spherical harmonic interpolation of HRTFs (see [18, Chapter 3.12.4]), which is one possible application of the dataset. Figure 1 shows the respective grids. For the circular grid measurements, the dummy head was mounted on a thin microphone stand, which again was fixed on the rotatable base plate of the VariSphear. When conducting the Lebedev grid measurements, the dummy head was fastened on the robot arm of the VariSphear. In combination with the rotatable base plate, this setup allowed for full 3D rotation of the head on a virtual sphere.

In total, nine HRIR datasets were captured. The circular grid was measured for five sound source distances (0.25 m, 0.5 m, 0.75 m, 1.00 m, 1.50 m) whereas the Lebedev grid was measured for only four distances (0.5 m, 0.75 m, 1.00 m, 1.50 m).

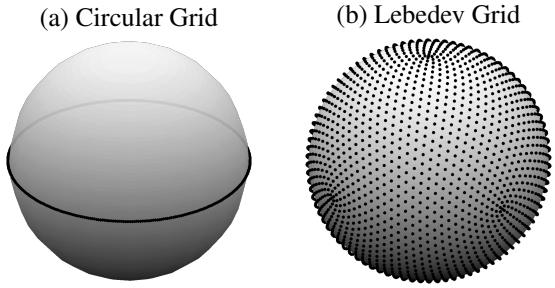

Fig. 1: Measured spatial sampling grids: Circular grid with steps of 1° in the horizontal plane (a) and Lebedev grid with 2702 points (b).

The closest distance was skipped here because the back of the robot arm would have touched the loudspeaker. A cross-line laser was used for precise positioning of the dummy head and the loudspeaker. For both setups, exact alignment of the head was checked for various sampling positions. The distance between the loudspeaker and the entrance of the dummy head's ear canal was determined accurately with a laser distance meter. This procedure was repeated for each new loudspeaker position. The acoustic center of the loudspeaker was always at ear level of the dummy head. Figure 2 exemplary shows the setup for the Lebedev grid measurements at a source distance of 0.5 m. Additionally, omnidirectional impulse responses were captured at the physical origin of the dummy head for all source distances with a Microtech Gefell M296S microphone. These reference measurements provided the basis for the magnitude and phase compensation of the loudspeaker, later on applied in post-processing (see section 2.2).

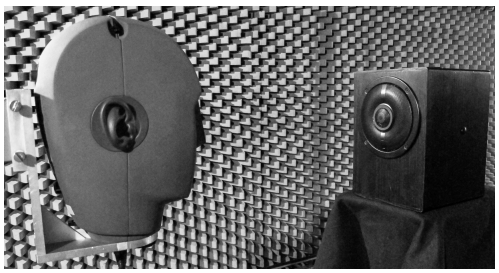

Fig. 2: Measurement setup in the anechoic chamber for the Lebedev grid measurement at a sound source distance of 0.5 m. The sound source was a Geithain RL906 two-way coaxial loudspeaker. The receiver was a Neumann KU100 dummy head, mounted on the VariSphear measurements system.

The excitation signal for all measurements was an emphasized sine sweep with +20 dB low shelf at 100 Hz. With 2^{19} samples at 48 kHz sampling rate, the sweep had a length of about 11 s, which allowed good robustness against background noise. The loudspeaker was driven at about -9 dB below its maximum permissible sound power level and the measurement peak level was always at about -6 dBFS. These settings yielded measurements with an overall SNR of about 90 dB. An RME Fireface UFX audio interface was used as AD / DA converter and microphone preamp. The whole measurement procedure was administered with the VariSphear software. Besides the motor control and impulse response capture modules, the software provided automatic error detection which checked every measured impulse response for noticeable variations with reference to previous measurements. This process ensured validity of all obtained impulse responses.

Eight kHz, LLC
Exhibit 2011
IPR2023-01024
Page 3

ISBN 987-3-9812830-7-5

Even though the measurements were conducted with great care, there are several shortcomings, which should be considered. First of all, the loudspeaker might violate the assumption of an acoustic point source in the proximal region (< 1.00 m). Moreover, there might be multiple reflections between loudspeaker and dummy head at close distances, resulting in HRTFs with increased ripple because of interferences. Another serious issue is the influence of the robot arm used for the Lebedev gird measurements. Whereas reflections at the arm are more or less negligible for frontal sound incidence ($\varphi = 0°$, $\delta = 0°$), the arm causes distinct shadowing effects for sound incidence from the rear ($\varphi = 180°$, $\delta = 0°$), which intensify with decreasing sound source distance (see section 2.3 for a more detailed explanation). Thus, the dataset (in particular the Lebedev grid data) should be considered as a valuable set for auralization purposes rather than as a basis for sensitive listening experiments.

2.2. Post-Processing

First, the raw measurement data were carefully truncated, windowed and transformed to the *miro* (measured impulse response object, [3]) format. Working with the MATLAB based miro data type allowed easy access to the datasets and convenient management of further processing. The two major aims of the post-processing were to achieve full range HRTF datasets by extending the low frequency range of the raw measurements and to compensate the influence of the loudspeaker by inverse FIR filtering. Most of the processing is based on the implementation and explanation from Bernschütz [3]. Thus, the following section focuses on the main aspects of the procedure and briefly outlines the processing steps and the technical motivation, whereas Bernschütz provides a more detailed explanation in his publication.

Adaptive Low Frequency Extension The low frequency range of raw HRTFs involves several inaccuracies. First of all, small loudspeakers, which are required for near-field measurements, typically fail to reproduce low frequencies (e.g. below 50 Hz) at adequate sound pressure levels. This leads to HRTFs with a distinct low frequency roll-off. Furthermore, particularly at low frequencies, the loudspeaker induces serious group delay. As a result, the HRIRs are more spread in time and thus more filter taps are required to cover the full audible frequency range. Another great problem is the sound field in the anechoic chamber below its cut-off frequency, where room modes and reflections arise. Because of this modal behavior, raw HRTF measurements show room and position dependent peaks and dips in the lower frequency range and therefore perform poorly when auralizing low frequency content. As a consequence, post-processing HRTFs at low frequencies is mostly necessary when full range datasets are required.

Replacing the low frequency range by an analytic expression is one well-suited approach for low frequency processing of HRTFs [3], [20]. For this purpose, Bernschütz [3] developed an algorithm for adaptive low frequency extension (ALFE), which we used for post-processing. The approach assumes that at frequencies below 400 Hz, pinna and ear canal of the dummy head hardly affect the HRTF and that the head itself has only minor influence on the sound field. According to that, it is reasonable to process HRTFs in order to obtain a flat magnitude response below a certain corner frequency less or equal than 400 Hz. Briefly speaking, the used ALFE-algorithm works as follows. The raw HRTF is high-pass filtered at a certain crossover frequency with a 24 dB/Oct Linkwitz-Riley filter and a matched low frequency extension (LFE) is attached, substituting the original low frequency component. This LFE corresponds to a time shifted Dirac delta function $\delta(n)$, adjusted in level according to the original low frequency component and low-pass filtered with the crossover filter. To match the phase slope of the filtered raw HRTF and the LFE around the crossover frequency, a first-order all-pass filter is applied. Since the algorithm is input-dependent, every raw HRTF as well as the reference measurements were processed separately. The crossover frequency was always set to 200 Hz whereas the cut-off frequency of the all-pass filter had to be adjusted per sound source distance. Figure 3 illustrates the described ALFE-processing in frequency domain. The improved HRTFs show a flat magnitude response below 200 Hz and, when examined in time domain, considerably less group delay.

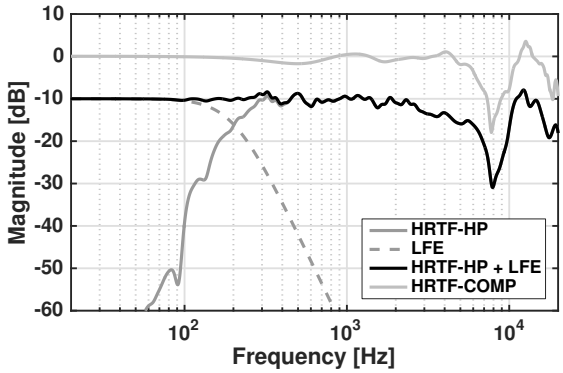

Fig. 3: HRTF post-processing applying adaptive low frequency extension (ALFE) and magnitude / phase compensation. *HRTF-HP*, 1/12-oct. smoothed high-pass filtered raw HRTF (left ear, $\varphi = 0°$, $\delta = 0°$, sound source distance = 1.50 m); *LFE*, low frequency extension - time shifted and low-pass filtered Dirac delta function; *HRTF-HP + LFE*, ALFE-processed HRTF - summed and phase-matched low and high frequency components; *HRTF-COMP*, final HRTF with ALFE-processing and magnitude / phase compensation.

Magnitude and Phase Compensation In a next step, magnitude and phase compensation were applied for further optimization. Therefore, we designed a specific compensation filter for each source distance, based on the ALFE-processed reference measurements. The respective compensation filter was implemented as a Hann-windowed FIR filter, basically describing the appropriately inverted frequency and phase response of the corresponding reference. Filtering all measurements removed further artifacts caused by the loudspeaker, like variations in magnitude response and remaining group delay. As a result of the compensation in time domain, the HRIRs could finally be truncated to 128 taps at 48 kHz sampling rate, while still maintaining the full spectral bandwidth. The length of the head and tail window was set appropriately in the miro files to ensure only negligible influence when windowing is applied. Figure 3 shows an example of a final HRTF in frequency domain.

Eight kHz, LLC
Exhibit 2011
IPR2023-01024
Page 4

ISBN 987-3-9812830-7-5

Final Processing In a last processing step, all datasets were slightly leveled so that the HRTFs for sound incidence from the front and from the rear approximate a magnitude of 0 dB at DC. This was more an aesthetic rather than a much-needed step since the deviations from 0 dB at DC were 1 dB at most. The leveling was not applied to the circular dataset with distance of 0.25 m because the peak level of HRIRs for lateral sound incidence would have exceeded 0 dBFS. However, even though all dataset were peak normalized, reconstructing the distance-dependent level differences is still possible based on the normalization factors listed in the miro metadata. Finally, the miro files were converted to the more common SOFA format [4] to provide usability for a wider user group.

2.3. Technical Evaluation

Near-field HRTFs usually have typical signal properties depending on the distance to the sound source, distinguishing them clearly from common far-field HRTFs. Brungart et al. [5, 8, 9], for example, presented a range of such near-field features in their extensive research on nearby sound sources. To check if our new HRTF set also shows the expected characteristics, we examined the final datasets carefully and extracted some of the main features. Moreover, we reviewed all data to check for any deficiencies caused by the measurement setup or post-processing. Please note that all of the following plots showing HRTF properties are based on the circular grid sets, mainly because these sets do not suffer from the influence of the robot arm and because the characteristics are mostly shown in the horizontal plane anyway.

One prominent feature of near-field HRTFs is the increase of ILDs (Interaural Level Differences) as a function of source proximity. According to Brungart [5], especially at sound source distances below 0.5 m, this rise of ILDs is dramatic. Hence, ILDs of near-field HRTFs show the typical increase as the source moves lateral to the head, which is basically caused by (frequency dependent) head shadowing effects. However, since these shadowing effects are much stronger at the contralateral ear and the magnitude at the ipsilateral ear increases simultaneously, the resulting ILDs are distinctly higher [5]. This effect can be easily observed in Figure 5(a), which shows the ILDs of our presented HRTF set for a sound source in the horizontal plane. Whereas the ILDs at the sound source distances 1.50 m, 1.00 m and 0.75 m are more or less similar, they start to increase at a distance of 0.5 m and escalate at the closest distance of 0.25 m. These ILDs up to about 23 dB might provide a relevant cue for distance perception in the proximal region. Off course, ILDs are frequency dependent; a fact also investigated in the context of near-field HRTFs by Brungart et al. [5].

Next, we examined the ITDs (Interaural Time Differences) of the presented HRTF set. Figure 5(b) displays the respective ITDs, calculated by the threshold onset method [21] including 10 times oversampling for more precise onset detection. As expected, the ITDs increase as the source moves lateral to the head and usually peak at about 90° and 270°. Both, the depicted ITDs and ILDs, show the familiar direction-dependent influence of the pinna and the head. However, unlike the ILDs, the ITDs are barely influenced by sound source distance, which is also in line with observations of Brungart et al. [5]. A closer look at Figure 5(b) reveals a slight increase of the time differences as distance decreases, leading to a maximum of about 742 μs at lateral positions and at a sound source distance of 0.25 m. This small rise appears because the length of the path from the ipsilateral to the contralateral side of the head increases as the source approaches. It goes without saying that ITDs and their behavior in the proximal region are also frequency dependent effects, as described more precisely in Brungart et al. [5].

Another prominent effect is the low-pass filtering character of proximal-region sources, meaning that sound sources are getting darker in timbre as they approach the head [5]. This effect is strongest for very close distances and sound sources at the front or rear. It appears because the ears are in the acoustic shadow zone of the head, which mainly damps higher frequencies. The spectral difference between the HRTF at 0.25 m and 1.50 m for frontal sound incidence, shown in Figure 4, demonstrates the described low-pass character. Again, it might be possible that this effect serves as a monaural cue for distance estimation in the proximal region.

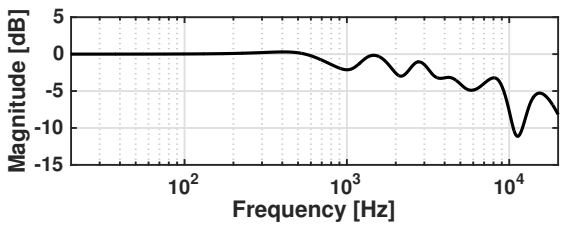

Fig. 4: Spectral difference (1/3-oct. smoothed) between the HRTFs at a source distance of 1.50 m and 0.25 m (left ear, $\varphi = 0°$, $\delta = 0°$). The result illustrates the low-pass filtering effect of proximal-region sources for frontal sound incidence.

In another analysis, we took a closer look at the influence of the robot arm. Therefore, we compared the Lebedev grid with the circular grid measurements at two distances (0.50 m, 1.50 m) and for sound incidence from the front ($\varphi = 0°$, $\delta = 0°$) and from the rear ($\varphi = 180°$, $\delta = 0°$) . By calculating the spectral differences between the respective Lebedev grid and circular grid HRTFs, the influence of the robot arm on the magnitude spectrum can be determined. As depicted in Figure 6, the robot arm only slightly affects the HRTFs for frontal sound incidence. The effect is more or less independent of sound source distance, mainly because the gap between the dummy head and the reflecting robot arm at the back of the head is alway the same. Overall, the reflections at the robot arm cause some minor interference artifacts in the final Lebedev grid HRTFs, starting at about 700 Hz. In the frequency range between 700 Hz and 20 kHz and at the distance of 0.50 m, the ripple has a mean of about 0.55 dB (SD = 0.59 dB) and a maximum absolute value of 2.25 dB at 7.2 kHz. For this particular case, the perceptual influence of the artifacts might be relatively small. For sound incidence from the rear, however, the robot arm causes strong shadowing effects and interferences, as shown in Figure 6. Here, at the distance of 0.50 m, the spectral difference has a mean of about 6.51 dB (SD = 4.61 dB) and maximal damping values of about 10 - 15 dB at frequencies above 10 kHz. At 1.50 m, especially the high frequency damping effect above 10 kHz is weaker, basically because the robot arm does not cover the tweeter

Eight kHz, LLC
Exhibit 2011
IPR2023-01024

ISBN 987-3-9812830-7-5

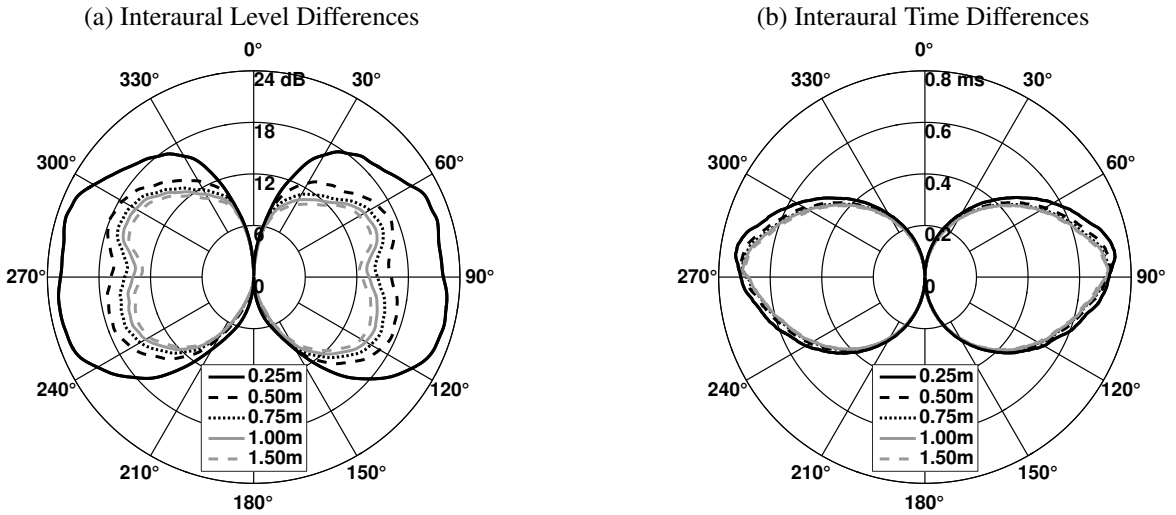

Fig. 5: Interaural Level Differences (a) and Interaural Time Differences (b) of the presented HRTF dataset. The angle represents the azimuth of the sound source (φ). The radius describes the magnitude of the level differences (in dB) or time differences (in ms).

of the loudspeaker. Nevertheless, these Lebedev grid HRTFs clearly suffer from the influence of the robot arm, regardless of the source distance. Both HRTFs lack high frequencies, which is plainly audible in auralizations, especially when compared to the corresponding circular grid HRTF.

Moreover, our signal analysis showed that the reflections between the loudspeaker and the dummy head only affect the post-processed circular grid HRTFs for a distance of 0.25 m. Truncating the HRIRs to 128 taps removed the reflections in the datasets for higher distances, simply because their delay exceeds the length of the HRIRs. Nevertheless, as already mentioned in the paragraph about shortcomings of the measurements in section 2.1, using the HRTF set for sensitive listening experiments should be carefully considered. However, regarding the key features of the presented HRTF set (thoroughly post-processed full range HRTFs, several distances in the near and far field, circular and full spherical grid), it is well suited for many auralization applications.

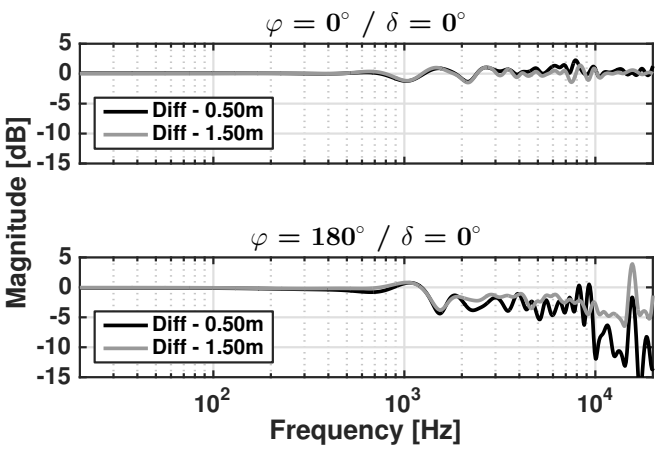

Fig. 6: Spectral differences (1/12-oct. smoothed) between circular grid and Lebedev grid HRTFs at a source distance of 0.50 m and 1.50 m and for sound incidence from the front ($\varphi = 0^\circ$, $\delta = 0^\circ$) and rear ($\varphi = 180^\circ$, $\delta = 0^\circ$). The results illustrate the minor influence of the robot arm on the Lebedev grid HRTFs for frontal sound incidence and its strong effect on the Lebedev grid HRTFs for sound incidence from the rear. In the latter case, the Lebedev grid HRTFs clearly lack high frequencies in comparison to the circular grid HRTFs.

3. Perceptual Evaluation

Based on the new HRTF set, we conducted several listening experiments within the context of auditory distance perception. In this paper, we present preliminary results for one part of this test series. Here, the basic task was to estimate auditory distance to a virtual sound source in dynamic binaural synthesis. The presented study served to investigate if the HRTFs can be applied to code distance and if appropriate distance estimation is still possible when natural level differences between the stimuli are missing. The latter is of particular interest, since the significant changes of binaural and monaural cues for a sound source in the near-field suggest that it is possible to distinguish distance (in the proximal region) even without the prominent factor level difference. Please note that the subjects had no previous training in distance estimation of nearby sound sources. Thus, they had to rely on their life experience in perceiving near-field sound sources.

3.1. Method

Participants Two females and 13 males aged between 21 and 28 years (M = 24.1 years, SD = 2.23) participated at this stage of the experiment. Most of them were students in media technology or electrical engineering. Thirteen participants already took part in previous listening experiments and thus were familiar with the binaural system. None of the subjects reported any hearing problems.

Setup The experiment took place in the anechoic chamber at TH Köln, which ensured a low background noise level of less than 20 dB(A). The experiment was implemented, controlled, and executed with the MATLAB-based software Scale [22], which also accessed the SoundScape Renderer [23] for binaural rendering. To acquire horizontal head movements, a Polhemus Fastrak head tracking system was used. Vertical or translational head movements were disregarded. The subjects entered their responses on a tablet computer (iPad). The audio signal was presented over AKG K-601 headphones. Headphone compensation was applied according to [3] in order to equalize the binaural chain.

Eight kHz, LLC
Exhibit 2011
IPR2023-01024

ISBN 987-3-9812830-7-5

Materials The anechoic test signal was a pink noise burst sequence with a burst length of 1500 ms (including 10 ms cosine-squared onset/offset ramps) and pauses of 500 ms. For the listening experiment, we used the circular grid measurements for all five distances from 0.25 m to 1.50 m. Per distance, we tested for three different sound incidence angles (φ = 30°, 150° and 270°). As already mentioned, we also wanted to test if appropriate distance estimation is still possible without natural level differences. Therefore, we prepared a second set of HRTFs, loudness-normalized with regard to the pink noise test stimuli according to ITU-R BS.1770. The playback level for the loudness-normalized conditions was at about 61 dB(A) Leq. For the non-normalized conditions, we assigned this playback level to a sound source distance of 1 m, resulting in a maximum playback level of about 79 dB(A) Leq for the closest distance of 0.25 m (φ = 270°).

Procedure As already mentioned above, there was no training session and no scale anchoring process. Informal pretests showed that training involved strong learning effects, especially for the normalized conditions: First, test persons could not immediately distinguish between distances, but when they were given feedback, they learned to differentiate based on spectral changes, varying ILDs and head movement. However, we wanted to know if distance perception in the near field works instantaneously without prior knowledge about the auditory scene. Therefore, we only gave a basic instruction about the general procedure and the rating scale.

The listening test was composed of two sessions. In the first session, subjects had to rate the normalized conditions, in the second session the non-normalized ones. Thus, the *normalization* order was blocked across participants. In each session, every participant had to rate the five measured *distances* (0.25 m, 0.50 m, 0.75 m, 1.00 m, 1.50 m) for three different source *azimuths* (φ = 30°, 150°, 270°). This resulted in a 5 × 3 × 2 within-subjects design.

Participants had to rate distance on a seven-point category scale ("very close", "close", "rather close", "medium", "rather distant", "distant", "very distant"); a scale that had been successfully used in earlier experiments [24]. It was allowed to rate interim values between the given categories. The procedure was as follows. For each trial, a user interface was displayed on the tablet computer containing five value faders ranging from "very close" to "very distant" (see Figure 7). The five faders corresponded to the five actual measured distances, thus the subjects had to rate multiple stimuli per trial. The source azimuth was the same for all distances (or faders) within a trial. By touching the respective fader, the participants were able to switch between the corresponding stimuli as often as required. Technically speaking, the HRTF filter-set switched when touching the fader while the noise sequence was played in a loop. The order of the faders per trial as well as the order of the trials itself were randomized. The procedure was repeated 10 times per azimuth, thus a full run consisted of 30 trials (with five distance ratings per trial). The listeners were encouraged to move their head during the estimation process in the form of (small) localization movements. However, they had to keep their front viewing direction because of the different source directions. In total, the test lasted for about one hour including the verbal instruction, one short break, and three post-experiment questions.

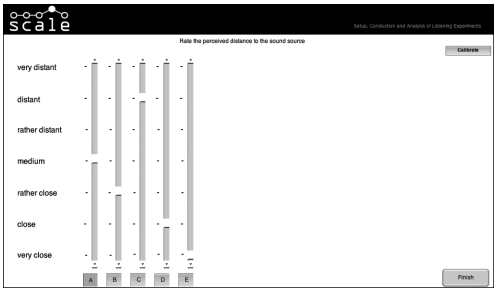

Fig. 7: User interface of the experiment. The left side displays the seven-point category scale. The five faders correspond to the five actual measured distances, randomly ordered for each trial.

3.2. Results

The following statistical analysis is based on the mean value per subject, thus the 10 trials per subject for each condition were averaged first. A 5 × 3 × 2 repeated measures ANOVA (distance, azimuth, normalization) with Greenhouse-Geisser (GG) correction [25] (for tests with more than one degree of freedom in the numerator, where GG is appropriate) was conducted. The ANOVA yielded a significant distance main effect ($F(4,56) = 71.91$, $p < .001$, $\eta_p^2 = .84$, $\epsilon = .41$) as well as significant interaction effects of distance × azimuth ($F(8,112) = 6.97$, $p = .004$, $\eta_p^2 = .33$, $\epsilon = .24$), distance × normalization ($F(4,56) = 34.78$, $p < .001$, $\eta_p^2 = .71$, $\epsilon = .30$) and distance × azimuth × normalization ($F(8,112) = 9.65$, $p = .001$, $\eta_p^2 = .41$, $\epsilon = .26$). Figure 8 presents the respective means of estimated distance per normalized (a) and non-normalized (b) conditions, averaged over subjects. The error bars display 95% within-subject confidence intervals [26], based on the error term of the distance main effect. The interaction effect of distance and azimuth is mainly caused by the variances for conditions with a source distance of 0.25 m (see Figure 8(a)). A repeated measures ANOVA without these conditions confirmed this: here, the interaction effect of distance and azimuth was not significant anymore ($p = .05$). More interesting seems to be the interaction effect of distance and normalization, which is why the mean plots in Figure 8 are split relative to the factor normalization. Presenting the results this way suggests that participants failed to distinguish distances for the normalized conditions (see Figure 8(a)). Without loudness-normalization, meaning with natural level differences between the stimuli, the results are as expected: the subjects rated according to the actual measured distances (see Figure 8(b)). A nested repeated measures ANOVA, each for the normalized and the non-normalized conditions, supported this assumption. Whereas there was no significant main effect of distance ($p = .71$) or azimuth ($p = .76$) for the normalized conditions, there was a strong distance main effect ($F(4,56) = 135.73$, $p < .001$, $\eta_p^2 = .91$, $\epsilon = .36$) for the non-normalized ones. These results indicate that the binaural and monaural cues characterizing sources in the near- and far-field do not influence distance estimation, even though the actual signal variations are huge in some cases (see section 2.3). The results were quite surprising, especially because we expected an effect of these cues, similar to Brungart et al. [8]. However, our results are rather in line with the findings from Shinn-Cunningham et al. [10] [11]. Hence, it appears that the participants rated

Eight kHz, LLC
Exhibit 2011
IPR2023-01024
Page 7

ISBN 987-3-9812830-7-5

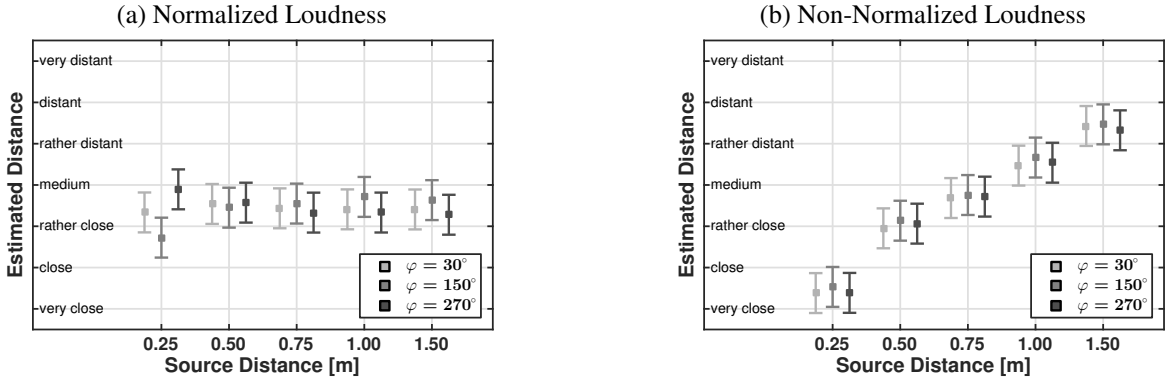

Fig. 8: Mean estimated distances for loudness-normalized (a) and non-normalized (b) conditions as a function of source distance (abscissa) and source azimuth (colors). The error bars denote 95% within-subject confidence intervals based on the respective main effect of distance.

distance mainly based on signal amplitude. Furthermore, a closer look at the results for normalized conditions with a distance of 0.25 m revealed large inter-subject differences. Some subjects correctly rated the proximal-region sources with their low-pass filtering character as close to the head (see section 2.3), whereas others assigned these conditions to very large distances, most likely because they interpreted the muffled sound as a result of high frequency energy dissipation. Overall, most subjects seemed not to have much experience in perception of nearby sound sources. Regarding the normalized conditions, the participants mostly stated that distance estimation was rather difficult and that they were very uncertain about the correct order of the stimuli.

4. Conclusion

Proper auralization of nearby sound sources requires near-field HRTFs with their specific features. In this paper, we presented a near-field HRTF set of a Neumann KU100 dummy head. The set contains post-processed impulse responses, measured according to two different spatial sampling grids: a Lebedev full spherical grid with 2702 points at four sound source distances (0.5 m, 0.75 m, 1.00 m, 1.50 m) and a circular grid with steps of 1° in the horizontal plane at five distances (0.25 m, 0.5 m, 0.75 m, 1.00 m, 1.50 m). After detailed explanations of the measurement setup and of the applied post-processing, we presented a technical evaluation of the final HRTF set and showed the typical (and expected) features of the near-field HRTFs. The final set served as the basis for a series of listening experiments within the context of auditory distance perception in anechoic environments. In the study presented in this paper, we investigated if the HRTFs can be applied to code distance and if appropriate distance estimation is still possible when natural level differences are missing. As expected, the preliminary results showed that distances can be distinguished when loudness-based distance cues exist, thus when the stimuli are not normalized in loudness. However, we observed that subjects could not estimate distances for loudness-normalized stimuli. These findings suggest that binaural cues do not affect distance estimation and vice versa, that auditory distance perception in anechoic environments mainly depends on loudness-based distance cues.

To go further into this issue, additional listening experiments need to be done. As already mentioned, the presented study is part of a larger test series concerning auditory distance perception of nearby sound sources. In an ongoing study, we investigate the influence of head tracking on distance estimation. Furthermore, it would be interesting to examine if a preceding training session influences distance estimation of nearby sound sources.

Apart from the listening experiments, which focus on specific research questions, our primary intention was to provide a freely available near-field HRTF dataset which is well suited for auralization purposes. Therefore, the set is available in the miro and SOFA format under a Creative Commons CC BY-SA 4.0 license and can be downloaded at: *http://audiogroup.web.th-koeln.de/ku100nfhrir.html*.

5. Acknowledgements

This work was funded by the German Federal Ministry of Education and Research (BMBF) under the support code 03FH014IX5-NarDasS. The authors thank all participants of the listening experiment. We thank Philipp Stade, Tim Lübeck and Patrick Pereira for their support during the experiments. The authors wish to thank Benjamin Bernschütz for his advice concerning the measurements and the post-processing.

6. References

[1] Algazi, V. R., Duda, R. O., and Thompson, D. M., "The CIPIC HRTF Database," in *IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics*, pp. 99–102, 2001.

[2] Gardner, W. G. and Keith, D. M., "HRTF measurements of a KEMAR," *J. Acoust. Soc. Am.*, 97(6), pp. 3907–3908, 1995.

[3] Bernschütz, B., "A Spherical Far Field HRIR / HRTF Compilation of the Neumann KU 100," in *Proceedings of the 39th DAGA*, pp. 592–595, 2013.

[4] Majdak, P., Iwaya, Y., Carpentier, T., Nicol, R., Parmentier, M., Roginska, A., Suzuki, Y., Watanabe, K., Wierstorf, H., Ziegelwanger, H., and Noisternig, M., "Spatially Oriented Format for Acoustics: A Data Exchange Format Representing Head-Related Transfer Functions," in *Proceedings of the 134th AES Convention, Rome, Italy*, pp. 1–11, 2013.

Eight kHz, LLC
Exhibit 2011
IPR2023-01024

[5] Brungart, D. S. and Rabinowitz, W. M., "Auditory localization of nearby sources. Head-related transfer functions," *J. Acoust. Soc. Am.*, 106(3), pp. 1465–1479, 1999.

[6] Stewart, G. W., "The Acoustic Shadow of a Rigid Sphere, with Certain Applications in Architectural Acoustics and Audition," *Phys. Rev.*, 33(6), pp. 467–479, 1911.

[7] Hartley, R. V. L. and Fry, T. C., "The Binaural Location of Pure Tones," *Phys. Rev.*, 18(6), pp. 431–442, 1921.

[8] Brungart, D. S., Durlach, N. I., and Rabinowitz, W. M., "Auditory localization of nearby sources. II. Localization of a broadband source," *J. Acoust. Soc. Am.*, 106(4), pp. 1956–1968, 1999.

[9] Brungart, D. S., "Auditory localization of nearby sources. III. Stimulus effects," *J. Acoust. Soc. Am.*, 106(6), pp. 3589–3602, 1999.

[10] Shinn-Cunningham, B. G., Santarelli, S., and Kopco, N., "Distance Perception of Nearby Sources in Reverberant and Anechoic Listening Conditions: Binaural vs. Monaural Cues," in *Poster presented at the 23rd MidWinter meeting of the Association for Research in Otolaryngology, St. Petersburg, Florida*, 2000.

[11] Shinn-Cunningham, B. G., "Localizing Sound in Rooms," in *Proceedings of the ACM SIGGRAPH and EUROGRAPHICS Campfire: Acoustic Rendering for Virtual Environments, Snowbird, Utah*, pp. 17–22, 2001.

[12] Kayser, H., Ewert, S. D., Anemüller, J., Rohdenburg, T., Hohmann, V., and Kollmeier, B., "Database of multichannel in-ear and behind-the-ear head-related and binaural room impulse responses," *EURASIP Journal on Advances in Signal Processing*, 2009, pp. 1–10, 2009.

[13] Wierstorf, H., Geier, M., Raake, A., and Spors, S., "A Free Database of Head-Related Impulse Response Measurements in the Horizontal Plane with Multiple Distances," in *Proceedings of the 130th AES Convention, London, UK*, pp. 1–4, 2011.

[14] Xie, B., Zhong, X., Yu, G., Guan, S., Rao, D., Liang, Z., and Zhang, C., "Report on Research Projects on Head-Related Transfer Functions and Virtual Auditory Displays in China," *J. Audio Eng. Soc.*, 61(5), pp. 314–326, 2013.

[15] Nishino, T., Hosoe, S., Takeda, K., and Itakura, F., "Measurement of the head related transfer function using the spark noise," in *Proceedings of 18th International Congress on Acoustics*, pp. 1437–1438, 2004.

[16] Hosoe, S., Nishino, T., Itou, K., and Takeda, K., "Development of Micro-Dodecahedral Loudspeaker for Measuring Head-Related Transfer Functions in The Proximal Region," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 5, pp. 329–332, 2006.

[17] Qu, T., Xiao, Z., Gong, M., Huang, Y., Li, X., and Wu, X., "Distance-Dependent Head-Related Transfer Functions Measured With High Spatial Resolution Using a Spark Gap," in *IEEE Transactions on Audio, Speech and Language Processing*, volume 17, pp. 1124–1132, 2009.

[18] Bernschütz, B., *Microphone Arrays and Sound Field Decomposition for Dynamic Binaural Recording*, Doctoral dissertation, TU Berlin, 2016.

[19] Bernschütz, B., Pörschmann, C., Spors, S., and Weinzierl, S., "Entwurf und Aufbau eines variablen sphärischen Mikrofonarrays für Forschungsanwendungen in Raumakustik und Virtual Audio (Design and Construction of a Variable Spherical Microphone Array for Research in Room Acoustics and Virtual Audio)," in *Proceedings of the 36th DAGA*, pp. 717–718, 2010.

[20] Xie, B., "On the low frequency characteristics of head-related transfer function," *Chinese Journal of Acoustics*, 28(2), pp. 116–128, 2009.

[21] Katz, B. F. G. and Noisternig, M., "A comparative study of interaural time delay estimation methods." *J. Acoust. Soc. Am.*, 135(6), pp. 3530–3540, 2014.

[22] Vazquez Giner, A., "Scale - Conducting Psychoacoustic Experiments with Dynamic Binaural Synthesis," in *Proceedings of the 41st DAGA*, pp. 1128–1130, 2015.

[23] Geier, M., Ahrens, J., and Spors, S., "The SoundScape Renderer: A Unified Spatial Audio Reproduction Framework for Arbitrary Rendering Methods," in *Proceedings of the 124th AES Convention, Amsterdam, The Netherlands*, pp. 1–6, 2008.

[24] Pörschmann, C. and Störig, C., "Investigations Into the Velocity and Distance Perception of Moving Sound Sources," *Acta Acustica united with Acustica*, 95(4), pp. 696–706, 2009.

[25] Greenhouse, S. W. and Geisser, S., "On Methods in the Analysis of Profile Data," *Psychometrika*, 24(2), pp. 885–891, 1959.

[26] Loftus, G. R. and Masson, M. E. J., "Using confidence intervals in within-subject designs," *Psychon. Bulletin & Review*, 1(4), pp. 476–490, 1994.

Eight kHz, LLC
Exhibit 2011
IPR2023-01024

ISBN 987-3-9812830-7-5

View publication stats