Audio Engineering Society

Convention Paper 10575

Presented at the 152nd Convention
2022 May, In-Person and Online

Acquisition of Continuous-Distance Near-Field Head-Related Transfer Functions on KEMAR Using Adaptive Filtering

Yuqing Li, Stephan Preihs, and Jürgen Peissig

Institut für Kommunikationstechnik, Leibniz Universität Hannover

Correspondence should be addressed to Yuqing Li (yuqing.li@ikt.uni-hannover.de)

ABSTRACT

Near-field Head-Related Transfer Functions (HRTFs) depend on both source direction (azimuth/elevation) and distance. The acquisition procedure for near-field HRTF data on a dense spatial grid is time-consuming and prone to measurement errors. Therefore, existing databases only cover a few discrete source distances. Coming from the fact that continuous-azimuth acquisition of HRTFs has been made possible by applying the Normalized Least Mean Square (NLMS) adaptive filtering method, in this work we applied the NLMS algorithm in measuring near-field HRTFs under continuous variation of source distance. We developed and validated a novel measurement setup that allows the acquisition of near-field HRTFs for source distances ranging from 20 to 120 cm with one recording. We then evaluated the measurement accuracy by analyzing the estimation error from the adaptive filtering algorithm and the key characteristics of the measured HRTFs associated with near-field binaural rendering.

1 Introduction

Virtual spatial audio can be rendered by convolving a mono signal with a pair of Head-Related Impulse Responses (HRIRs) or, equivalently in the frequency domain, by multiplying the Fourier transform of the signal with the Head-Related Transfer Functions (HRTFs). HRTFs describe the acoustical alterations caused by a listener's head, torso and pinnae on incident sound waves from a point source in space. These alterations include sound scattering, diffraction, and absorption. As a result, interaural differences and spectral coloration are introduced into the incoming sound waves, giving the listener an impression of the spatiality of the sound and helping them locate the sound source.

HRTFs are dependent on the relative position between the sound source and the listener. When the source is far from the listener, HRTFs only vary with the direction of the source, and the term "far-field HRTF" applies; when the source comes close to the head, i.e. within about 1 meter from the head center, the so-called "near-field HRTF" shows distance-dependent variations. The dominant characteristics of near-field HRTFs on the horizontal plane has been summarized in some literature [1] [2]:

- In the near-field, as the source distance decreases, there is an increase in the Interaural Level Difference (ILD) across all frequencies but particularly at low frequencies.
- The Interaural Time Difference (ITD) is relatively distance-independent compared to the ILD.
- With decreasing source distance, the magnitude of the near-field HRTF increases faster at low fre-

quencies than at high frequencies. Therefore, approaching sound sources appear to be low-pass filtered.

- High-frequency features in the ipsilateral HRTF can be shifted laterally as the source approaches the head, due to the difference between the source-ear orientation and the source-head orientation for a nearby source. This phenomenon is named "the parallax effect".

Capturing and reproducing these characteristics is important for creating realistic virtual auditory events in the vicinity of the listener. Because all the acoustical cues for binaural hearing are included in HRTFs, the physical and perceptual quality of HRTF-based binaural reproduction in a virtual auditory display largely depends on the accuracy of the HRTFs. Only reproducing point sources at fixed positions is not sufficient to satisfy the expanding demand on the functionality of virtual reality applications. Especially when clustered sources or dynamic auditory events are considered, the spatial resolution of the HRTF dataset must be adequately high to avoid audible artifacts during binaural rendering.

Although various types of measurement apparatus and systems have been proposed, the measurement of near-field HRTFs is still a sophisticated and tedious procedure because the relative position between the sound source and the listener must be changed many times to cover different directions and distances. As a result, there are only a small number of measured near-field HRTF databases, e.g. [2] and [3], and the distance resolution of these databases is considerably low. If a pair of HRTFs is desired at a position between the measured points, they must be calculated using interpolation and extrapolation methods. Alternatively, high-resolution near-field HRTFs can be obtained from numerical simulations or structural-model-based adaptation of far-field HRTFs. For example, Salvador et al. [4] calculated near-field HRTFs for four head models from 10 to 100 cm with 1 cm spacing using the boundary element method; Kan et al. [5] utilized a "Distance Variation Function (DVF)" to simulate near-field HRTFs from far-field HRTF measurements based on a rigid sphere sound scattering model; Arend and Pörschmann [6] proposed an approach that combines the DVF method and spherical harmonics (SH)-based processing to synthesize near-field HRTFs by spectral equalization and spatial upsampling of far-field datasets.

Since the continuous-azimuth HRTF measurement method using the Normalized Least Mean Square (NLMS) adaptive filtering algorithm was introduced by Enzner [7], it has proven to be an effective workaround for fast acquisition of HRTFs with high angular resolution. In this approach, the source direction is continuously changed during the measurement with either a rotating sound source [7] or subject [8], and HRTFs for all directions can be measured with only one recording. Since HRTFs are assumed to be continuous functions in 3D space (i.e. in both direction and distance) [9], it is possible to apply the same concept to the acquisition of multi-distance HRTFs by continuously varying the source distance during measurement. In this work, we propose a method for continuous-distance near-field HRTF measurement using the NLMS algorithm.

The paper is structured as follows. Section 2 introduces the application of the NLMS algorithm to the measurement of continuous-distance near-field HRTFs; Section 3 presents a specially designed measurement system for the proposed approach; objective evaluation of the results is given in Section 4; Section 5 summarizes the entire work.

2 NLMS Adaptive Filtering for Continuous HRTF Acquisition

By adapting the NLMS algorithm used by Enzner [7], we obtain the following linear convolution model for estimating time-varying HRIR filters with continuously varying source distance (as illustrated in Figure 1), where the ear signals can be formulated as:

$$y_i(k) = \sum_{\kappa=0}^{N-1} x(k-\kappa) h_i(\kappa, d_k) + n_i(k) \qquad (1)$$

Where $i \in \{1,2\}$ is the ear index (left/right), d_k the source distance at discrete time k, x the input signal which can be the loudspeaker signal or a reference recording, h_i the HRIR filter to be estimated, and n_i the additional measurement noise. The linear convolution model is only valid under the assumption that the time corresponding to the rate of change in the HRIR with source distance is significantly longer than the filter length N [7].

The most recent N input samples into the system can be denoted in vector form as:

$$\mathbf{x}(k) = (x(k), x(k-1), \ldots, x(k-N+1))^T \qquad (2)$$

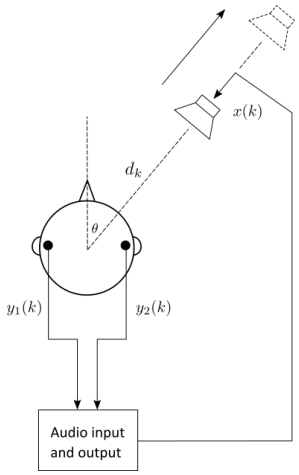

Fig. 1: Illustration of the measurement setup based on continuous distance variation.

Similarly, the HRIR coefficients can be expressed as:

$$\mathbf{h}_i(d_k) = (h_i(0,d_k), h_i(1,d_k), \ldots, h_i(N-1,d_k))^T \quad (3)$$

The estimation error $e_i(k)$ from (1) is the sample-based deviation between the measured output and the estimated output:

$$e_i(k) = y_i(k) - \widehat{\mathbf{h}}_i^T(d_k)\,\mathbf{x}(k) \quad (4)$$

And an estimation of the HRIR filter at distance d_k is recursively calculated by:

$$\widehat{\mathbf{h}}_i(d_{k+1}) = \widehat{\mathbf{h}}_i(d_k) + \mu_0 \frac{e_i(k)\mathbf{x}(k)}{\|\mathbf{x}(k)\|_2^2} \quad (5)$$

Therefore, the filter coefficients are constantly adapted as the sample number k increases, so that the mean-square error $E[e_i^2(k)]$ is minimized.

Since we are only interested in near-field HRTFs, the source distance d_k is constrained by a lower and an upper bound, denoted by D_n and D_f, respectively. The behavior of the NLMS algorithm is largely influenced by the stepsize μ_0 and the movement speed v of the sound source, where $v = (D_f - D_n)/T_{total}$, T_{total} is the total time for the sound source to move the distance between D_n and D_f. For a faster movement, a larger stepsize μ_0 is needed to ensure a sufficient convergence speed of the algorithm.

3 Measurement and Processing of Continuous-Distance Near-Field HRTFs of KEMAR

3.1 Measurement apparatus

A special measurement apparatus was developed for the proposed approach. A motorized camera slider was used to carry the sound source and move it continuously at a constant speed along a straight trajectory. The slider originally came with a USB controller which can only be controlled manually. However, for the purpose of our work, we must ensure that the exact position information is known at any time in the recorded signal. Therefore, the slider's motor was instead driven by custom-made microcontroller-based hardware and software, so that the movement can be triggered and controlled by a User Datagram Protocol (UDP) packet sent from MATLAB. The UDP packet specifies all the parameters that define a movement: the start position in centimeters (relative to one end of the slider), the number of rounds, the moving distance in centimeters, and the movement time in seconds. To verify that the movement speed is constant, we used the Optitrack motion tracking system with Prime 17W cameras to collect position data from the loudspeaker as it moves, triggered by the same UDP packet used for the HRTF measurement. The deviation of the actual position from the assumed one (i.e. calculated from the defined speed and elapsed time) is always less than 0.6 cm.

The measurement took place in an anechoic room at Institut für Kommunikationstechnik, Leibniz Universität Hannover. A G.R.A.S. KEMAR was used, whose built-in mircophones (G.R.A.S. 40HH 1/2" Lownoise Microphone System) served as receivers for the binaural signals. The KEMAR was placed on a digitally controlled turntable for taking measurements in different azimuth directions.

The slider was supported at each end by a tripod carefully adjusted to the same height. The sound source used was a miniature loudspeaker with a 3D printed enclosure. The frequency response of the loudspeaker is shown in Figure 2. The frontal board of the loudspeaker has a radius of 4 cm. To minimize unwanted reflections, the surfaces of the tripods, the front of the slider and the loudspeaker were acoustically treated with sound-absorbing materials. Figure 3 shows a photo of the measurement setup.

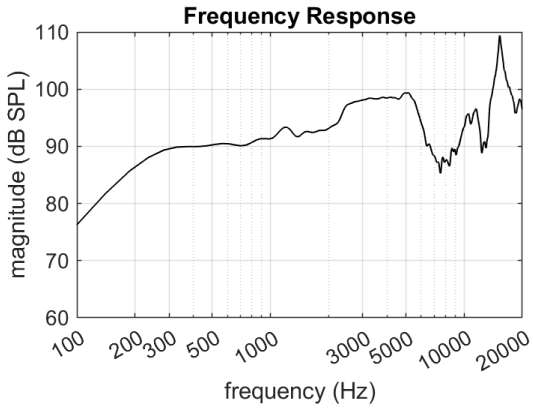

Fig. 2: Frequency response of the loudspeaker (measured under $\leq 5\%$ Total Harmonic Distortion between 300 and 20k Hz).

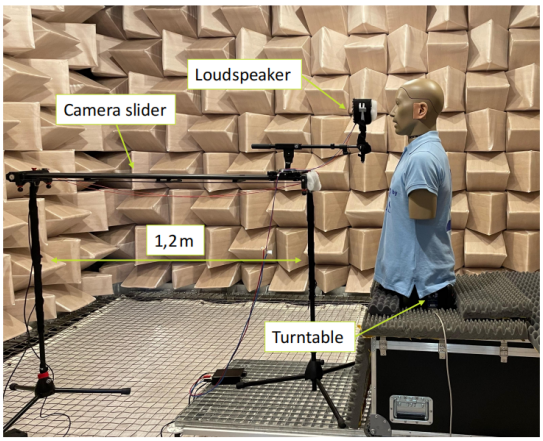

Fig. 3: Measurement setup for continuous-distance HRTF acquisition.

The smallest and largest source distance allowed by the appratus were 20 cm and 120 cm, respectively. The azimuthal resolution of measurement was 5° from 0° to 355°, where 0° corresponds to the front, 90° to the right and 180° to the back of the KEMAR. Measurements were taken in the horizontal plane only. The movement of the slider, the rotation of the turntable, the audio playback and recording were all controlled by a computer located in a room next to the anechoic room.

We used a perfect-sweep-based excitation signal because it is robust against non-linear behavior of the measurement system and allows fast tracking for NLMS estimation [10]. The excitation signal was constructed by concatenating identical perfect sweeps with a length of 512 samples. There was an "overhead" of 1 second at the beginning and end of the excitation signal for providing some buffer time for the NLMS algorithm, i.e. the sound began playing before the movement started and continued after the movement stops. In addition, a magnitude envelope was applied to the excitation signal according to the inverse distance law of sound pressure to compensate for distance-related attenuation, so that sufficient sound energy reaches the microphones even at the largest source distances. The total moving time of the loudspeaker from one end to the other was 30.034 seconds.

To improve the signal-to-noise ratio (SNR) and speed up the procedure, one recording was taken in each measurement as the loudspeaker moved back and forth twice, and the aforementioned excitation signal was repeated four times in one take (a forward movement is when the source moves away from the KEMAR). When the loudspeaker moved backwards, the signal was mirrored in time. At each azimuth angle, 4 recordings were taken, resulting in 8 averages for each direction (forward/backward) of movement. Each measurement took about 130 seconds. Therefore, the recording of the binaural signals for all azimuthal directions can ideally be completed in about 10.4 hours. The actual measurement time was about 16 hours, including pauses between takes and time for some retakes due to inevitable interrupting factors.

As with traditional HRTF measurements, a reference measurement is required to eliminate the acoustical effects from the rest part of the measurement system, e.g. the transfer function of the loudspeaker. Reference recordings were taken with the G.R.A.S 40HL 1/2" measurement microphone system. The KEMAR was removed and the microphone was placed with its diaphragm in the same position as the head center of the KEMAR. The frequency responses of all the microphones used can be considered as flat (variations within ± 2 dB) from 10 Hz to 16 kHz.

To evaluate the level of measurement noise, especially that induced by the mechanics of the moving units, binaural recordings were also taken without the excitation signal. The results show that an SNR of 40-60 dB can be achieved in the ipsilateral ear at 500-10000 Hz.

3.2 HRTF extraction with the NLMS algorithm, selection of stepsize and objective analysis of the estimation accuracy

Based on the nature of the NLMS algorithm, there are two ways to extract HRIR data from the raw recordings.

Approach 1: Time-varying impulse responses can be extracted separately from the reference recording and the binaural recordings, with the original loudspeaker signal as input $\mathbf{x}(k)$ for both cases. After that, an HRTF can be computed by taking a division between the FFTs of a binaural impulse response and a reference impulse response corresponding to the same spatial position.

Approach 2: Since the time alignment between the source positions and the recorded signals are identical for all measurements, a reference recording can be used as input $\mathbf{x}(k)$ to the NLMS algorithm and HRIRs can be calculated directly.

Intuitively, Approach 2 only involves time-domain processing and is therefore computationally simpler. Rounding errors from additional complex division operations can also be avoided. Therefore, Approach 2 was used in this work for the calculation of continuous-distance HRTFs. Before the computation, the sensitivity differences between the reference microphone and the binaural microphones were corrected.

Under the assumption that the distance-dependent variation in the HRIRs is neglectable within the time corresponding to N = 512 samples, the highest HRIR distance resolution that can be achieved with the current setup is 0.036 cm. Considering the limitation of resource and time, a distance resolution of 1 cm was chosen for HRIR extraction.

The stepsize μ_0 determines the speed of convergence and is usually set between 0 and 2 [11]. A large μ_0 leads to faster convergence, but possibly larger estimation errors. To determine a suitable range for μ_0, an objective metric of the estimation accuracy must be defined and tested for different μ_0 values. Since the actual HRIRs are not known, the estimation error $e_i(k)$ is the only available indicator for the quality of the calculated HRIRs. A small $E[e_i^2(k)]$ does not necessarily suggest a small system distance, because the output signal varies with time in our case. Therefore, the estimation quality was quantified by the Normalized Mean Square Error (NMSE), defined as follows:

$$NMSE = \frac{E[e_i^2(k)]}{E[y_i^2(k)]} \tag{6}$$

To evaluate the estimation quality at different distances, an NMSE value was calculated for each frame in the binaural signals corresponding to a stored HRIR. Figure 4 shows the NMSE values calculated from both forward and backward recordings for 3 different choices of μ_0: 0.75, 1 and 1.25, at 4 azimuths. As expected, regardless of the movement direction, $\mu_0 = 0.75$ gives the lowest NMSE after reaching steady-state, but shows a relatively slow convergence. $\mu_0 = 1.25$ enables fast convergence especially for non-lateral directions, but the steady-state NMSE is higher. For backward recordings, an increase in the NMSE after reaching steady-state is observed, possibly suggesting increased distance-dependent variation in the HRIRs due to additional scattering effects as the source moves closer to the head, which were not present earlier when the source was far away. The fact that ipsilateral NMSEs are more affected by the value of μ_0 supports this assumption. Forward recordings generally produce lower steady-state NMSE, whereas the estimation error for the first a few HRIRs is large, and outliers appear for 0° azimuth at some distances within 50 cm. Because we are more interested in the behavior of HRTFs at close distances, the first 50 HRIRs (corresponding to 20-69 cm) computed from the forward measurements were replaced by the last 50 HRIRs from the backward measurements. As the improvement in steady-state NMSE is small (< 2 dB) when changing from $\mu_0 = 1$ to $\mu_0 = 0.75$, yet for $\mu_0 = 1$ in lateral directions, lower NMSEs can be achieved for small source distances in the backward measurement, the stepsize was set to 1.

4 Results

4.1 Validation of the continuous measurement

We first validated the proposed approach by comparing the continuous-distance near-field HRTFs with those from existing databases, which were measured in static setups. Disparity in HRTFs will inevitably arise from different measurement apparatus and signal processing approaches used by different laboratories. Therefore, we are concerned about whether the amount of deviation of the continuous-distance HRTFs from those in other databases is acceptable. A commonly used metric to measure the similarity between two HRTFs

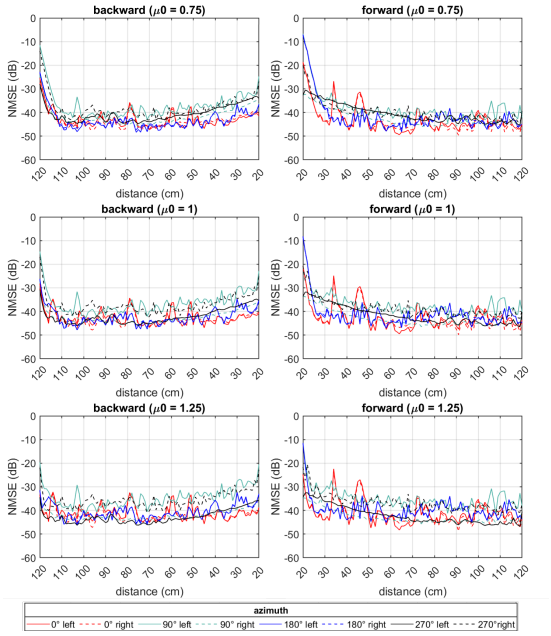

Fig. 4: NMSE values calculated with different μ_0.

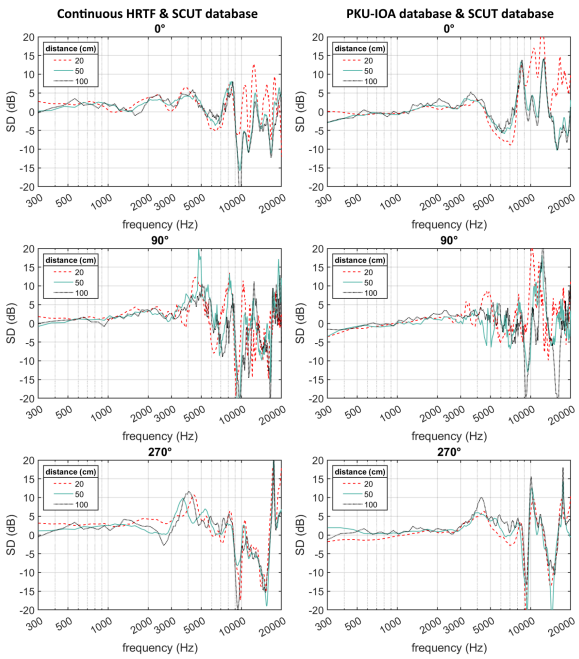

Fig. 5: First column: SD between the measured HRTFs and HRTFs from the SCUT database. Second column: SD between HRTFs from the PKU-IOA database and the SCUT database. Left-ear HRTFs were used for calculation.

is spectral distortion (SD). It is the euclidean distance between the two spectra integrated in the entire frequency range [12][13][14]. In our case, however, the frequency-dependent spectral difference is more of interest. The frequency-dependent SD between two HRTFs is defined as follows (distance and azimuth notations are omitted for readability):

$$SD(f) = 20\log_{10}\frac{|H_1(f)|}{|H_2(f)|} \tag{7}$$

We calculated SD at different distances and azimuths between our measured HRTFs and the SCUT database [2] , and between the PKU-IOA database [15] and the SCUT database as reference. Both are near-field HRTF databases measured on KEMAR. Before the calculation of SD, HRTFs at each distance were normalized. Results for 3 distances and 3 azimuths are plotted in Figure 5.

The two columns in Figure 5 are generally similar to each other in terms of the pattern and scope of variation of frequency-dependent SD. Hardly any distance-related SD variation can be observed, since the curves representing different distances are clustered in all of the plots. The SD remains relatively low and constant at frequencies up to 3000 Hz. As frequency increases, larger SD values are observed at some frequencies, accompanied by more fluctuations in magnitude, especially in the contralateral ear (90°). The most dramatic peaks at some high frequencies are observed in both columns. Therefore, we conclude that the proposed approach is capable of yielding results with comparable accuracy to existing near-field HRTF databases.

4.2 Objective evaluation on time- and frequency-domain features related to near-field binaural rendering

The HRIR magnitude is plotted in Figure 6 as a function of time and distance. Reflections from the frontal surface of the loudspeaker appear as secondary peaks in the ipsilateral (270°) HRIRs. The delay time of these secondary peaks increases linearly with distance, and the magnitude decreases with distance. Considering that in a real-world listening scenario, the sound source always has a physical dimension and reflections can

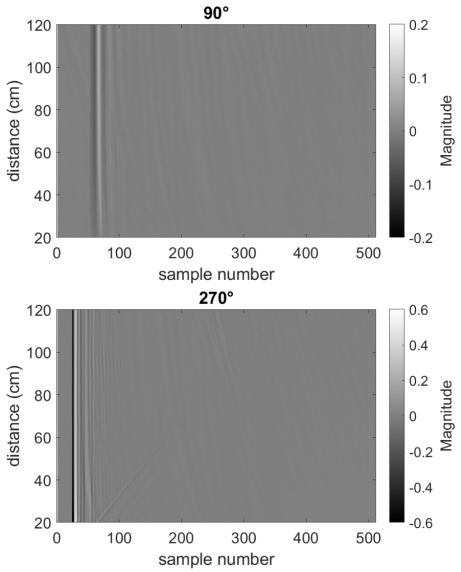

Fig. 6: Left-ear HRIR magnitude at contralateral and ipsilateral directions for all measurement distances (low-pass filtered, passband frequency: 16 kHz).

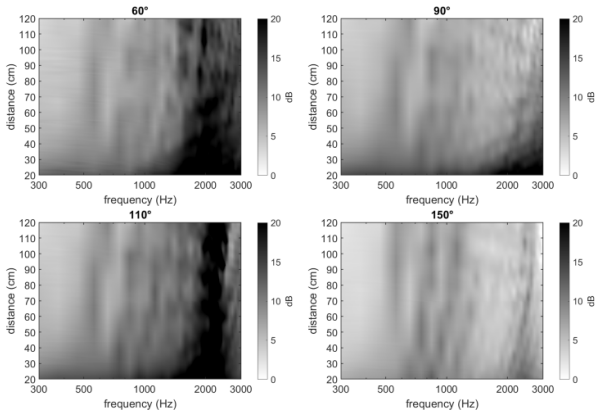

Fig. 7: Interaural Level Difference (ILD) below 3 kHz, calculated by taking a division between the right ear HRTF magnitude and the left ear HRTF magnitude.

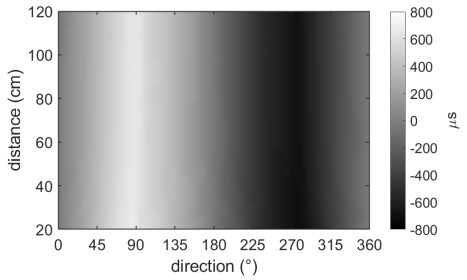

Fig. 8: Interaural Time Difference (ITD).

not be avoided, we do not regard the effect of these reflections to be detrimental to binaural playback through headphones in a virtual auditory display. The secondary peaks are not visible in the contralateral HRIRs because the reflected sound has limited energy and would be largely scattered in both space and time before reaching the contralateral ear. The other line-like structures which occur later in time and show an opposite distance-delay relationship (i.e. less delay at larger distances) are probably caused by reflecting objects near the more distant end of the slider.

Another conspicuous feature is that at close source distances (i.e. < 50 cm), as the source approaches the head, the peaks of the contralateral HRIRs are shifted later in time and the peak magnitude reduces. The explanation for this is that sound waves from a close source are more attenuated and delayed at the contralateral ear because of enhanced head shadowing [1], i.e. longer propagation path and more surface scattering. On the other hand, the peak magnitude of ipsilateral HRIRs increases with decreasing distance.

According to Brungart and Rabinowitz [1], the most perceptually relevant cue for near-field localization is the increase in low-frequency (< 3000 Hz) Interaural Level Difference (ILD) with decreasing distance. It can be observed in Figure 7 for source distances less than 50 cm. The closer the sound source to the interaural axis, the stronger this effect. Beyond 50 cm, the ILD is smaller and shows less variation with the change in distance. A similar observation was made by Kan et al. [16]. The reduced ILD around 2000 - 3000 Hz for source distances over 40 cm at 90° compared to other directions is caused by the formation of the acoustic bright spot [1], which becomes more prominent when the source moves away from the head.

Along with the ILD, the Interaural Time Difference (ITD) also serves as an important localization cue. The ITD values were calculated by detection of the -10 dB onset threshold [17]. As shown in Figure 8, the ITD hardly demonstrates any distance-dependent variation, as expected.

4.3 Analysis of the spatial variation of near-field HRTFs

The data obtained with the proposed approach offer the possibility for understanding the (pseudo-)continuous spatial variation of near-field HRTFs. The continuous HRTF magnitude is plotted against direction and distance in Figure 9. This form of representation is called the Spatial Frequency Response Surface (SFRS) [18] and provides an intuitive impression of the distribution of HRTF energy at different spatial locations for a given frequency. Across all frequencies, a reduction in HRTF energy is observed in the contralateral ear, which becomes enhanced with decreasing distance and increasing frequency, except at 3000 Hz where the HRTF magnitude at all distances and directions shows a peak due to the quarter-wavelength resonance of the ear canal, according to [1]. Most of this distance-dependent magnitude reduction arises when the source distance is less than 40-50 cm. The azimuth range in which the reduction occurs also becomes wider as the source approaches.

The acoustic bright spot is most evident at 1500 Hz as a local peak near 90°, which is then replaced by several lobes at higher frequencies. Some high frequency features in the ipsilateral ear are shifted towards the direction of the interaural axis as the source distance decreases due to the parallax effect. These observations are consistent with the generally recognized near-field HRTF features as mentioned in Section 1.

5 Summary

In this paper, we presented an algorithm and introduced a novel measurement setup for the acquisition of continuous-distance HRTFs in the near-field using the NLMS adaptive filtering approach. The proposed method was validated by objective evaluation of the computational accuracy and analysis of the near-field HRTF characteristics. With this method, a near-field HRTF dataset can be measured with high resolution in both direction and distance in a short time. Furthermore, the measurement results can be used for various research objectives related to near-field binaural rendering, such as analysis on the spatial variation of near-field HRTFs and validation of different HRTF acquisition approaches.

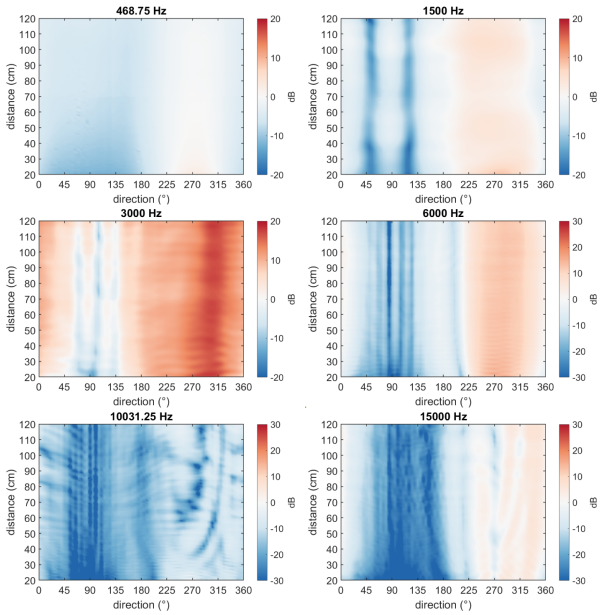

Fig. 9: Spatial Frequency Response Surfaces of left-ear HRTF magnitude.

References

[1] Brungart, D. S. and Rabinowitz, W. M., "Auditory localization of nearby sources. Head-related transfer functions," *The Journal of the Acoustical Society of America*, 106(3), pp. 1465–1479, 1999, doi:10.1121/1.427180.

[2] Yu, G.-Z., Xie, B.-S., , and Rao, D., "Characteristics of Near-field Head-Related Transfer Function for KEMAR," in *AES 40th International Conference: Spatial Audio: Sense the Sound of Space*, 2010.

[3] Arend, J., Neidhardt, A., and Pörschmann, C., "Measurement and Perceptual Evaluation of a Spherical Near-Field HRTF Set," in *29th Tonmeistertagung - VDT International Convention*, pp. 356–363, 2016.

[4] Salvador, C. D., Sakamoto, S., Treviño, J., and Suzuki, Y., in *Dataset of Near-Distance Head-Related Transfer Functions Calculated Using the Boundary element Method*, 2018.

[5] Kan, A., Jin, C., and van Schaik, A., "Distance Variation Function for Simulation of Near-Field

Virtual Auditory Space," in *2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*, volume 5, pp. 325–328, 2006, doi:10.1109/ICASSP.2006.1661278.

[6] Arend, J. and Pörschmann, C., "Synthesis of Near-Field HRTFs by Directional Equalization of Far-Field Datasets," in *45th DAGA*, pp. 1454–1457, 2019.

[7] Enzner, G., "Analysis and Optimal Control of LMS-type Adaptive Diltering for Continuous-azimuth Acquisition of Head Related Impulse Responses," in *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 393–396, 2008, doi:10.1109/ICASSP.2008.4517629.

[8] He, J., Ranjan, R., Gan, W.-S., Chaudhary, N., Duy, N., and Gupta, R., "Fast Continuous Measurement of HRTFs with Unconstrained Head Movements for 3D Audio," *Journal of the Audio Engineering Society*, 66(11), pp. 884–900, 2018, doi:10.17743/jaes.2018.0050.

[9] Xie, B., *Head-Related Transfer Function and Virtual Auditory Display*, J. Ross Publishing, 2 edition, 2013.

[10] Antweiler, C., Telle, A., Vary, P., and Enzner, G., "Perfect-sweep NLMS for time-variant acoustic system identification," in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 517–520, 2012, doi:10.1109/ICASSP.2012.6287930.

[11] Li, S., Tobbala, A., and Peissig, J., "Towards Mobile 3D HRTF Measurement," in *148th Convention of the Audio Engineering Society*, 2020.

[12] Tommasini, F., Ramos, O., Hüg, M., and Bermejo, F., "Usage of Spectral Distortion for Objective Evaluation of Personalized HRTF in the Median Plane," *International Journal of Acoustics and Vibration*, 20, pp. 81–89, 2015, doi:10.20855/ijav.2015.20.2371.

[13] Richter, J.-G., *Fast Measurement of Individual Head-Related Transfer Functions*, Ph.D. thesis, RWTH Aachen University, 2019.

[14] Jo, H., Park, Y., and Park, Y.-s., "Analysis of individual differences in Head-Related Transfer Functions by spectral distortion," in *2009 ICCAS-SICE*, pp. 1769–1772, 2009.

[15] 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," *IEEE Transactions on Audio, Speech, and Language Processing*, 17(6), pp. 1124 – 1132, 2009, doi:10.1109/TASL.2009.2020532.

[16] Kan, A., Jin, C., and van Schaik, A., "A psychophysical evaluation of near-field head-related transfer functions synthesized using a distance variation function," *The Journal of the Acoustical Society of America*, 125(4), pp. 2233–2242, 2009, doi:10.1121/1.3081395.

[17] Katz, B. F. G. and Noisternig, M., "A comparative study of interaural time delay estimation methods," *The Journal of the Acoustical Society of America*, 135(6), pp. 3530–3540, 2014, doi:10.1121/1.4875714.

[18] Cheng, C. I., *Visualization, Measurement, and Interpolation of Head-Related Transfer Functions (HRTF's) With Applications in Electro-Acoustic Music*, Ph.D. thesis, University of Michigan, 2001.