Parameter Analysis for Range Extrapolation of Head-Related Transfer Functions using Virtual Local Wave Field Synthesis

Fiete Winter, Sascha Spors

Institute of Communications Engineering, University of Rostock, R.-Wagner-Str. 31 (H8), 18119 Rostock, Germany

Email: {fiete.winter, sasha.spors}@uni-rostock.de

Introduction

Binaural synthesis utilizing head-related transfer functions (HRTFs) is a common approach to auralize virtual acoustic sources. HRTFs represent the acoustic free field transmission path from the source to the outer ears. They capture the acoustic characteristics of the outer ears which are exploited by the human auditory system in order to deduce spatial information. HRTFs differ amongst individuals due to varying anatomy. They are additionally depending on the head/body-orientation and position with respect to the source. HRTFs are typically measured in anechoic environments. For a virtual acoustic scene, left and right ear drum signals are rendered by filtering an anechoic signal of a virtual sound source with the left and right ear HRTFs.

In order to enable arbitrary positioning of the virtual sound source, a (densely) sampled grid of HRTFs is necessary. It is obvious that the measurement effort for the required HRTF dataset would be considerable. Hence, typical datasets are available for various source directions (on a circle or on a sphere) but only for a few distances up to 3 meters [1]. The characteristics of the HRTFs are generally assumed to be invariant with regard to source distances exceeding this threshold [2]. However, HRTFs of nearby sound sources significantly depend on the distance [2].

In the past several approaches have been proposed to extrapolate an HRTF for a desired distance from an available HRTF dataset. Two of them expand the HRTFs into surface spherical harmonics in order to perform extrapolation in the spherical harmonics domain [3, 4]. Other methods interpret the available HRTF measurements as a virtual loudspeaker array, which has be to driven according to the desired source position. While some of these techniques [5, 6] are based on Higher Order Ambisonics (HOA), a numerically stable and computationally efficient method uses Wave Field Synthesis (WFS) [7].

The latter method was further refined in [8] using Local Wave Field Synthesis (LWFS) [9], which utilizes a distribution of focused sources as so-called *virtual secondary sources*. These are placed more densely and nearer to the listener than the original HRTF dataset. This paper analyses, in how far the number of the virtual secondary sources and their distance to the listener influences the accuracy of this extrapolation method. This is done by comparing the extrapolated HRTFs with measured ones with regard to important characteristics, i.e. the magnitude spectrum and the average Interaural Level Difference (ILD).

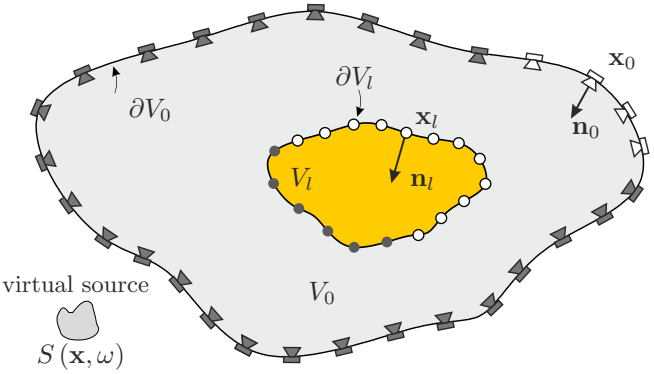

Figure 1: In Local Wave Field Synthesis the desired sound field $S(\mathbf{x}, \omega)$ is reproduced inside the local listening area V_l (yellow shade) with the virtual secondary source distribution (dots) on its boundary ∂V_l. Active virtual secondary sources and loudspeakers are shaded dark.

Local Wave Field Synthesis

The goal of LWFS is to reproduce a desired sound field $S(\mathbf{x}, \omega)$ of a virtual source within in so-called *local listening area* V_l (see Fig. 1). This can be achieved with a surrounding distribution of loudspeakers (also termed secondary sources) located at ∂V_0. The (3D) free-field Green's function $G_0(\mathbf{x} - \mathbf{x}_0, \omega)$ characterizes the sound field emitted by one loudspeaker with its position $\mathbf{x}_0 \in \partial V_0$. The synthesized sound field

$$P(\mathbf{x}, \omega) = \oint_{\partial V_0} D_0(\mathbf{x}_0, \omega)\, G_0(\mathbf{x} - \mathbf{x}_0, \omega)\, \mathrm{d}A_0\,, \qquad (1)$$

is then given as the superposition of all loudspeakers which are individually driven by $D_0(\mathbf{x}_0, \omega)$. The boundary element $\mathrm{d}A_0$ is suitably chosen for integration. In LWFS, the loudspeakers are used to establish a distribution of virtual secondary sources, which has to be driven like a real loudspeaker setup. The loudspeakers' driving function is therefore given as

$$D_0(\mathbf{x}_0, \omega) = \oint_{\partial V_l} D_l(\mathbf{x}_l, \omega)\, D_{\mathrm{fs}}(\mathbf{x}_0 - \mathbf{x}_l, \omega)\, \mathrm{d}A_l\,. \qquad (2)$$

The driving function $D_{\mathrm{fs}}(\mathbf{x}_0 - \mathbf{x}_l, \omega)$ creates the impression of one virtual secondary source at the focus point $\mathbf{x}_l \in \partial V_l$. Each virtual secondary source is driven by the traditional WFS driving function

$$D_l(\mathbf{x}_l, \omega) = -2\, a_l(\mathbf{x}_l)\, \frac{\partial}{\partial \mathbf{n}_l} S(\mathbf{x}, \omega)\,, \qquad (3)$$

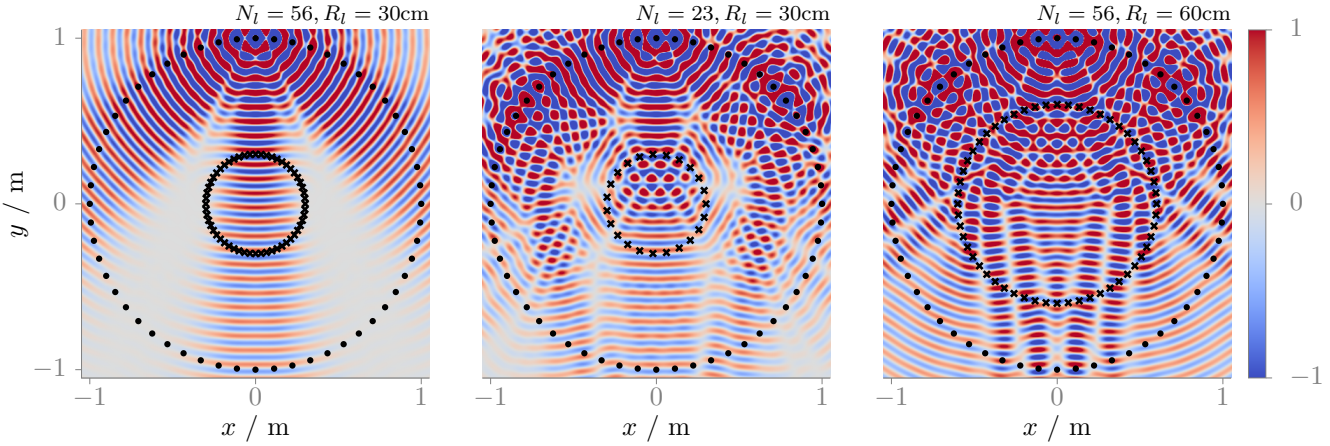

Figure 2: The plots show the real part of the reproduced sound field $P(\mathbf{x}, \omega)$ of a virtual, monochromatic ($f = 4$kHz) point source at $\mathbf{x}_{\text{ps}} = [0, 1.5, 0]^T$m. A circular array of 56 Loudspeakers (black dots) with a radius of 1m is used for reproduction. The number N_l and the radius R_l of the virtual secondary sources (black crosses) is varied among the three plots. All sound fields are normalized to their respective values at $[0, 0, 0]^T$.

where the directional gradient $\frac{\partial}{\partial \mathbf{n}_l}$ is defined as scalar product of the boundary's inward normal vector $\mathbf{n}_l$ and the gradient $\nabla S(\mathbf{x}, \omega)$ evaluated at $\mathbf{x} = \mathbf{x}_l$. The secondary source selection criterion $a_l(\mathbf{x}_l)$ ensures that only those virtual secondary sources are active where the propagation direction of the virtual source $S(\mathbf{x}, \omega)$ at the position $\mathbf{x}_l$ has a positive component in direction of the normal vector $\mathbf{n}_l$.

While Eq. (1) and (2) assume continuous boundaries, practical setups introduce discretization of the (virtual) secondary source distribution. Hence, both integrals migrate to sums over discrete positions $\mathbf{x}_0$ and $\mathbf{x}_l$, respectively. The coarser the spatial resolution of either of the distributions, the more spatial aliasing is contributed to the reproduced sound field [10]. Figure 2 exemplarily shows that a larger local listening area demands additional virtual secondary sources in order to avoid aliasing. The reproduction accuracy of LWFS is furthermore bounded by the number of loudspeakers.

Range Extrapolation of HRTFs

The HRTF dataset is interpreted as emerging from a virtual loudspeaker setup. In order compute an extrapolated HRTF, these loudspeakers are driven by LWFS. Typical HRTF measurements are performed for source position on a circle or on a sphere. In this paper, the HRTFs are assumed to be measured on a regular, circular grid in the horizontal plane for the ease of illustration (see Fig. 3). This is covered by the theory of 2.5D WFS [10, Sec. 4]. The proposed method can nevertheless be extended to three-dimensional datasets, since LWFS allows for any convexly shaped loudspeaker setup. The circular, local listening area is centered around the listeners head.

The head-above-torso orientation of the listener states an additional degree of freedom for the HRTF measurements and can be considered by the extrapolation method straightforwardly [8]. It is however not in the focus of this publication and remains fixed for convenience.

Hence, the extrapolated HRTF for the apparent sound source position $\mathbf{x}_{\text{ps}}$ is given as

$$\tilde{H}_{\{\text{L,R}\}}(\mathbf{x}_{\text{ps}}, \omega) = \sum_{n=0}^{N_0-1} D_0(\mathbf{x}_{0,n}, \omega)\, H_{\{\text{L,R}\}}(\mathbf{x}_{0,n}, \omega). \quad (4)$$

This equation is derived from eq. (1) by replacing the Green's function with the measured $H_{\{\text{L,R}\}}(\mathbf{x}_{0,n}, \omega)$ and by discretizing the integral to N measurement positions $\mathbf{x}_{0,n} = R_0[\cos(\alpha_{0,n}), \sin(\alpha_{0,n}), 0]^T$. The positions are equally spaced with respect to their azimuth angle, i.e. $\alpha_{0,n} = 2\pi n / N_0$. The loudspeakers' driving function

$$D_0(\mathbf{x}_{0,n}, \omega) = \sum_{m=0}^{N_l-1} D_l(\mathbf{x}_{l,m}, \omega)\, D_{\text{fs}}(\mathbf{x}_{0,n} - \mathbf{x}_{l,m}, \omega) \quad (5)$$

is derived analogously by discretizing eq. (2). The position of the N_l (equally distributed) virtual secondary sources are denoted by $\mathbf{x}_{l,m} = R_l[\cos(\alpha_{l,m}), \sin(\alpha_{l,m})]^T$. The 2.5D driving function for a focused source at $\mathbf{x}_l$ is given as [8, eq. (4)]

$$D_{\text{fs}}(\mathbf{x}_0 - \mathbf{x}_l, \omega) = \frac{1}{2\pi} a_{\text{fs}}(\mathbf{x}_0 - \mathbf{x}_l) \sqrt{-\text{j}\frac{\omega}{c}} \times g_{2.5\text{D}}(\mathbf{x}_0) \frac{(\mathbf{x}_0 - \mathbf{x}_l)^T \mathbf{n}_l}{|\mathbf{x}_0 - \mathbf{x}_l|^{3/2}} \text{e}^{j\frac{\omega}{c}|\mathbf{x}_0 - \mathbf{x}_l|}, \quad (6)$$

where the secondary selection criterion for a focused source reads

$$a_{\text{fs}}(\mathbf{x}_0 - \mathbf{x}_l) = \begin{cases} 1 & \text{, if } \mathbf{n}_l^T(\mathbf{x}_0 - \mathbf{x}_l) < 0 \\ 0 & \text{, otherwise} \end{cases} \quad (7)$$

and $\mathbf{n}_l = -[\cos(\alpha_l), \sin(\alpha_l), 0]^T$. The virtual secondary sources are driven by traditional WFS [8, eq. (6)]

$$D_l(\mathbf{x}_l, \omega) = \frac{1}{2\pi} a_l(\mathbf{x}_l) \sqrt{\text{j}\frac{\omega}{c}} \times g_{2.5\text{D}}(\mathbf{x}_l) \frac{(\mathbf{x}_l - \mathbf{x}_{\text{ps}})^T \mathbf{n}_l}{|\mathbf{x}_l - \mathbf{x}_{\text{ps}}|^{3/2}} \text{e}^{-j\frac{\omega}{c}|\mathbf{x}_l - \mathbf{x}_{\text{ps}}|} \quad (8)$$

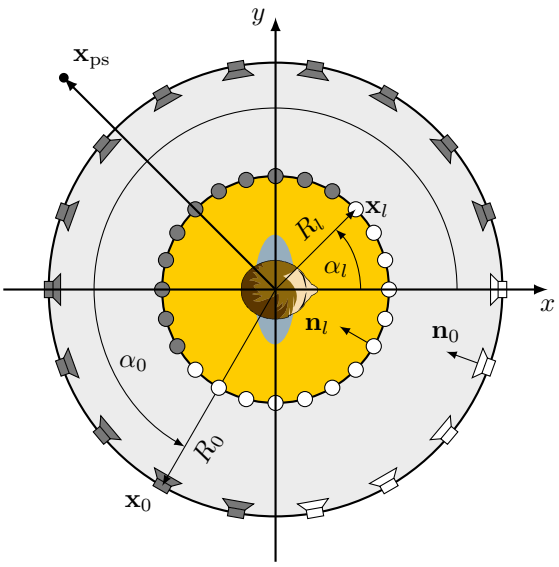

Figure 3: The loudspeaker symbols illustrate the virtual loudspeaker setup, i.e. the measured HRTF dataset. The virtual secondary source distribution is depicted by the bullets. Active virtual secondary sources and loudspeakers are shaded dark.

according to the desired extrapolated position of the point source $\mathbf{x}_{\text{ps}}$. The selection criterion for the virtual secondary source reads

$$a_l\left(\mathbf{x}_l\right) = \begin{cases} 1 & \text{, if } \mathbf{n}_l^T(\mathbf{x}_l - \mathbf{x}_{\text{ps}}) > 0 \\ 0 & \text{, otherwise.} \end{cases} \tag{9}$$

As a beneficial side effect of this criterion, inactive virtual secondary sources can be discarded from the computation. This reduces the complexity of the extrapolation approach. The number of active virtual secondary sources is denoted by N_l'.

A well known artefact of 2.5D reproduction is a systematic mismatch of the amplitude decay between the desired and the reproduced sound field. Hence, these deviations will also occur in the ILDs of the extrapolated HRTFs, especially for lateral sources [8]. $g_{2.5\text{D}}(\mathbf{x}_{\{0,l\}}) = \sqrt{2\pi|\mathbf{x}_{\{0,l\}} - \mathbf{x}_{\text{ref}}|}$ [11, eq. (4)] is a geometry dependent correction factor, which ensures an approximately correct amplitude at the reference position $\mathbf{x}_{\text{ref}}$. This parameter can be chosen for each ear separately in order to compensate the amplitude deviations.

Evaluation

The evaluation is performed by comparing measured HRTFs with their extrapolated pendant. For this purpose the HRTF dataset of [12] is used. It was captured with a Knowles Electronics Manikin for Acoustic Research (KEMAR), type 45BA, for source positions of different distance ($\{0.5, 1, 2, 3\}$m) and azimuth from -180° to 179° with a azimuthal resolution of 1°. In the experiment, HRTFs with a distance of 1m are used to extrapolate HRTFs of 3m distance. The reference point $\mathbf{x}_{\text{ref}}$ is placed at center of the listeners head $[0, 0, 0]^T$.

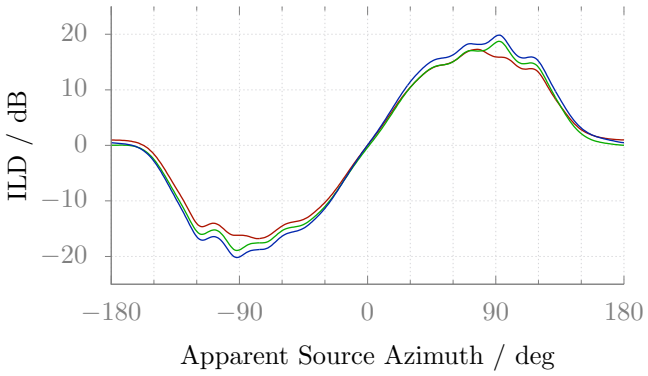

Figure 4: The plot shows average ILD of the HRTFs for a sound source with a distance of 3m. The ILD of measured HRTFs (red) is shown as a reference. The green and the blue line show ILDs of the extrapolation with $\mathbf{x}_{\text{ref}} \approx [-0.5, \pm 7.9, -3]^T$cm and $\mathbf{x}_{\text{ref}} \approx [0, 0, 0]^T$, respectively.

The effects of the parameters N_l' (number of active virtual secondary sources) and R_l (radius of the local listening area) on the magnitude spectrum of the extrapolated HRTFs are illustrated in Fig. 5: For a fixed size of the local listening area, aliasing artefacts are clearly visible at high frequencies, if an insufficient number of virtual secondary sources is used. Increasing N_l' beyond 60 does not further improve the result. The effects of R_l can be split into two aspects: Obviously artefacts occur when the virtual secondary sources are too near or "inside" the listeners head. Larger radii cause aliasing at high frequencies, since more virtual secondary sources are necessary for the resulting size of the listening area.

It has been outlined in the last section, that the systematic amplitude deviations of 2.5D sound reproduction distort the ILD. This can be seen in Fig. 4, when comparing the extrapolated result with the reference. Setting the reference position for each ear individually to $\mathbf{x}_{\text{ref}} \approx [-0.5, \pm 7.9, -3]^T$cm (position of the ears used in [13]) slightly improves the result. The amplitude distortion is not completely compensated, though.

Conclusion

This work analysed the influence of different parameters on the accuracy of an HRTF extrapolation technique using Local Wave Field Synthesis. In the experiments the usage of 60 active virtual secondary source with a distance of 30cm to the head yield the best accuracy. In general, the results state a tradeoff between computional complexity and achievable precision. They further establish a connection between the physical accuracy of sound field reproduced by LWFS and the resulting ear signals. This motivates further investigation on the perceptual attributes of differently parameterized LWFS setups.

It has been shown, that a sensible choice of the reference position for both ears can improve the extrapolation. It is however a non-trivial task to adjust the reference position in order to achieve an optimal compensation of the 2.5D related amplitude deviations. The perceptual impact of this parameter is also unclear and its investigation remains as future work.

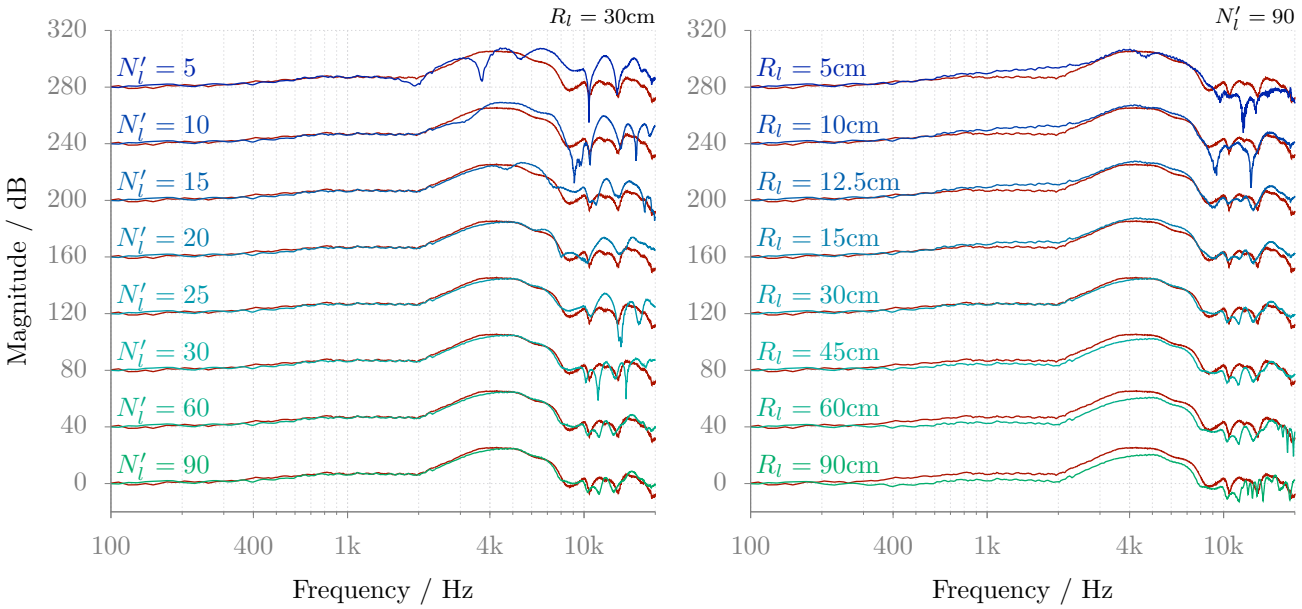

Figure 5: The plots show the magnitude spectra of the extrapolated left ear HRTF for a sound source with an apparent azimuth of 45° and a distance of 3m using different parametrizations for LWFSs. The measured HRTF (red) for this source position is shown as a reference in both plots. All spectra were normalized to their respective value at 100 Hz and are shifted about 40 dB each for a better visualization.

Acknowledgements

This research has been supported by EU FET grant Two!Ears, ICT-618075.

References

[1] B. Gardner and K. Martin, "HRTF Measurements of a KEMAR Dummy-Head Microphone," Tech. Rep. 280, MIT Media Lab, 1994.

[2] D. S. Brungart and W. M. Rabinowitz, "Auditory localization of nearby sources. Head-related transfer functions," *J. Acoust. Soc. Am.*, vol. 106, no. 3, pp. 1465–1479, 1999.

[3] R. Duraiswami, D. N. Zotkin, and N. A. Gumerov, "Interpolation and range extrapolation of HRTFs," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, (Montreal, Canada), 2004.

[4] W. Zhang, T. D. Abhayapala, R. A. Kennedy, and R. Duraiswami, "Modal expansion of HRTFs: Continuous representation in frequency-range-angle," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, (Taipei, Japan), Apr. 2009.

[5] M. Noisternig, A. Sontacchi, T. Musil, and R. Holdrich, "A 3D Ambisonic Based Binaural Sound Reproduction System," in *Proc. of 24th Intl. Aud. Eng. Soc. Conf. on Multichannel Audio*, (Banff, Canada), 2003.

[6] D. Menzies and M. Al-Akaidi, "Nearfield binaural synthesis and ambisonics," *J. Acoust. Soc. Am.*, vol. 121, no. 3, pp. 1559–1563, 2007.

[7] S. Spors and J. Ahrens, "Efficient Range Extrapolation of Head-Related Impulse Responses by Wave Field Synthesis Techniques," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, (Prague, Czech Republic), 2011.

[8] S. Spors and J. Ahrens, "Interpolation and Range Extrapolation of Head-Related Transfer Functions using Virtual Local Sound Field Synthesis," in *Proc. of 130th Aud. Eng. Soc. Conv.*, (London, UK), 2011.

[9] S. Spors and J. Ahrens, "Local Sound Field Synthesis by Virtual Secondary Sources," in *Proc. of 40th Intl. Aud. Eng. Soc. Conf. on Spatial Audio*, (Tokyo, Japan), 2010.

[10] S. Spors, R. Rabenstein, and J. Ahrens, "The theory of Wave Field Synthesis revisited," in *Proc. of 124th Aud. Eng. Soc. Conv.*, (Amsterdam, The Netherlands), 2008.

[11] S. Spors, H. Wierstorf, M. Geier, and J. Ahrens, "Physical and Perceptual Properties of Focused Sources in Wave Field Synthesis," in *Proc. of 127th Aud. Eng. Soc. Conv.*, (New York, USA), 2009.

[12] H. Wierstorf, M. Geier, and S. Spors, "A Free Database of Head Related Impulse Response Measurements in the Horizontal Plane with Multiple Distances," in *Proc. of 130th Aud. Eng. Soc. Conv.*, 2011.

[13] V. R. Algazi, C. Avendano, and R. O. Duda, "Elevation localization and head-related transfer function analysis at low frequencies," *J. Acoust. Soc. Am.*, vol. 109, no. 3, pp. 1110–1122, 2001.