Are spectral elevation cues in head-related transfer functions distance-independent?

Simone Spagnol
IUAV - University of Venice

ABSTRACT

Since its title, this paper addresses one of the still open questions in sound localization: is our own perception of the elevation of a sound source affected by the distance of the source itself? The problem is addressed through the analysis of a recently published distance-dependent head-related transfer function (HRTF) database, which includes the responses of a single subject on a spatial grid spanning 14 elevation angles, 72 azimuth angles, and 8 distances comprised between 20 and 160 cm. Different HRTFs sharing the same angular coordinates are compared through spectral distortion and notch frequency deviation measurements. Results indicate that, even though the independence of spectral elevation cues from distance of the source can be assumed for the majority of all possible source directions, near-field HRTFs for sources close to the contralateral ear or around the horizontal plane in the ipsilateral side of the head are significantly affected by distance-dependent pinna reflections and spectral modifications.

1. BACKGROUND

It is undisputed that vertical localization is possible thanks to the presence of the pinnae [1]. Even though localization in any plane involves pinna cavities of both ears [2], determination of the perceived elevation angle of a sound source in the median plane is essentially a monaural process [3]. The external ear plays an important role by introducing peaks and notches in the high-frequency spectrum of the head-related transfer function (HRTF), whose center frequency, amplitude, and bandwidth greatly depend on the elevation angle of the sound source [4] and to a remarkably minor extent on azimuth [5]. Following two historical theories of localization, the pinna can be seen both as a filter in the frequency domain [6] and a delay-and-add reflection system in the time domain [7] as long as typical pinna reflection delays for elevation angles, clearly detectable by the human hearing apparatus [8], are seen to produce spectral notches in the high-frequency range.

Nevertheless, the relative importance of major peaks and notches in elevation perception has been disputed over the past years. A recent study [9] showed how a parametric HRTF recomposed using only the first, omnidirectional peak in the HRTF spectrum (corresponding to Shaw's mode 1 [10]) coupled with the first two notches yields almost the same localization accuracy as the corresponding measured HRTF. Additional evidence in support of the lowest-frequency notches' relevance is given in [11], which states that the threshold for perceiving a shift in the central frequency of a spectral notch is consistent with the localization blur (i.e., the angular threshold for detecting changes in the direction of a sound source) on the median plane. Also, in [12] the authors judge increasing frontal elevation apparently cued by the increasing central frequency of a notch, and determine two different peak/notch patterns for representing the above and behind directions.

In general, hence, both pinna peaks and notches seem to play a primary function in vertical localization of a sound source, [1] even though it is difficult without extensive psychoacoustic evaluations to ascertain how importantly these features work as spatial cues. It has to be highlighted, however, that vertical localization bears little resolution compared with horizontal localization [13]. For the sake of record, the localization blur along the median plane was found to be never less than $4°$, reaching a much larger threshold ($\approx 17°$) for unfamiliar speech sounds, as opposed to a localization blur of approximately $1° - 2°$ in the horizontal plane for a vast class of sounds [6]. Such a poor resolution is motivated by two basic observations:

- the need of high-frequency content (above $4-5$ kHz) for accurate vertical localization [12, 14];
- the theoretically nonexistent interaural differences between the signals arriving at the left and right ear in the median plane.

Still, distance estimation of a sound source (see [15] for a comprehensive review on the topic) is even more troublesome than elevation detection. At a first level, when no other cue is available, sound intensity is the first variable that is taken into account: the weaker the intensity, the farther the source should be perceived. Under anechoic conditions, sound intensity reduction with increasing distance can be predicted through the inverse square law: intensity of an omnidirectional sound source will decay of approximately 6 dB for each doubling distance [16]. Still, a distant blast and a whisper at few centimeters from the ear could produce the same sound pressure level at the

[1] In this context, it is important to point out that both peaks and notches in the high-frequency range are perceptually detectable as long as their amplitude and bandwidth are sufficiently marked [11], which is the case for most measured HRTFs.

eardrum. Having a certain familiarity with the involved sound is thus a second fundamental requirement [17].

However, the apparent distance of a sound source is systematically underestimated in an anechoic environment [18]. On the other hand, if the environment is reverberant, additional information can be given by the proportion of reflected to direct energy, the so-called *R/D ratio*, which functions as a stronger cue for distance than intensity: a sensation of changing distance occurs if the overall intensity is constant but the R/D ratio is altered [16]. Furthermore, distance-dependent spectral effects also have a role in everyday environments: higher frequencies are increasingly attenuated with distance due to air absorption effects.

Literature on source direction perception generally lies its foundations on a fundamental assumption, i.e. the sound source is sufficiently far from the listener. In particular, both previously discussed elevation cues as well as azimuth cues such as interaural time and level differences (ITD and ILD) are distance-independent when the source is in the so-called *far field* (approximately more than 1.5 m from the center of the head) where sound waves reaching the listener can be assumed to be plane. On the other hand, when the source is in the *near field* some of the HRTF features exhibit a clear dependence on distance. By gradually approaching the sound source to the listener's head in the near field, it was observed that low-frequency gain is emphasized; ITD slightly increases; and ILD dramatically increases across the whole spectrum for lateral sources [19]. In this paper, Brungart and Rabinowitz drew the following conclusions:

- ITD is roughly independent of distance even when the source is close;
- low-frequency ILDs are the dominant auditory distance cues in the near field;
- elevation cues are not correlated to distance-dependent features in the near field.

It should be then clear that ILD-related information needs to be considered in the near field, where dependence on distance cannot be approximated by a simple inverse square law. However, in [19] the last conclusion is supported just by graphical evidence on a limited number of HRTFs. Specifically, it is shown with the support of a single figure that the major features of the HRTFs at three distinct elevations and three distances are considerably more consistent across distance than across elevation. The authors argue that *"if this result generalizes to all elevations, it would imply that elevation cues are roughly independent of distance and that the same mechanisms that mediate elevation perception in the distal region* (i.e. the far field) *may also be used in the proximal region* (i.e. the near field)*"* but, to the best of my knowledge, this hypothesis has never been verified in the following literature. My aim in this paper is thus to investigate more deeply – through the analysis of a new distance-dependent HRTF database [20] – whether Brungart's claim on the rough independence between elevation cues and distance in the near field is well-grounded.

2. ANALYSIS OF DISTANCE-DEPENDENT HRTFS

Typically, HRTFs are measured by presenting a sound stimulus at several different spatial locations lying on the surface of a sphere centered in the subject's head, hence at one single distance (typically 1 m or farther). Most public HRTF databases, such as CIPIC [21] and LISTEN [22], follow this standard. Measuring HRTFs at closer distances introduces technical difficulties because a common loudspeaker cannot simulate an acoustic point source in the near field, and the sound source should be as compact as possible in order to avoid sound reflections from the loudspeaker back into the microphones [23].

Recently, Qu *et al.* [20] successfully overcame the problem by using a specialized spark gap as an appropriate acoustic point source (from the viewpoints of frequency response, signal-to-noise ratio, directivity, power attenuation and stability) to collect a spatially dense set of HRTFs of a KEMAR manikin. The database, that was updated and made available in June 2012,[2] includes the responses at both the left and right ears for 72 different azimuth angles, 14 different elevation angles, and 8 different distances ranging from 20 to 160 cm from the center of the manikin's head, totalling 6344 HRTFs. The so obtained HRTFs were seen to be comparable to the well known and widely used KEMAR HRTFs included in the CIPIC database. The following analysis will be based on the whole set of right-ear HRTFs.

2.1 Spectral distortion

In order to have a first quantification of the difference between HRTFs for various distances at fixed azimuth (θ) and elevation (ϕ) angles, an error measure widely used in recent literature [20, 24] is introduced: spectral distortion

$$SD(H,\tilde{H}) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(20\log_{10}\frac{|H(f_i)|}{|\tilde{H}(f_i)|}\right)^2} \quad \text{[dB]}, \tag{1}$$

where H and $\tilde{H}$ are the responses to be compared and N is the number of available points in the considered frequency range, that I choose to be $R_1 = [500, 16000]$ Hz in order to include all the possible elevation cues. In the following analysis the reference HRTF, $\tilde{H}$, will always be the response for the farthest distance ($d_8 = 160$ cm) approximating the far field response, while H will be the HRTF for one of the closest distances sharing the same angular coordinates (θ_k, ϕ_k).

The analysis first requires the normalization of the responses in order to eliminate sound intensity cues. This is simply accomplished by dividing each HRTF by its mean magnitude in R_1,

$$\hat{H}(f,\theta,\phi,d) = \frac{H(f,\theta,\phi,d)}{\sum_{i=1,N}|H(f_i,\theta,\phi,d)|} \times N. \tag{2}$$

The resulting normalized HRTFs for fixed angular coordinates are approximately aligned in magnitude, as is the case for $(\theta,\phi) = (0°, 0°)$ in Figure 1.

[2] `http://www.cis.pku.edu.cn/auditory/Staff/Dr.Qu.files/Qu-HRTF-Database.html`

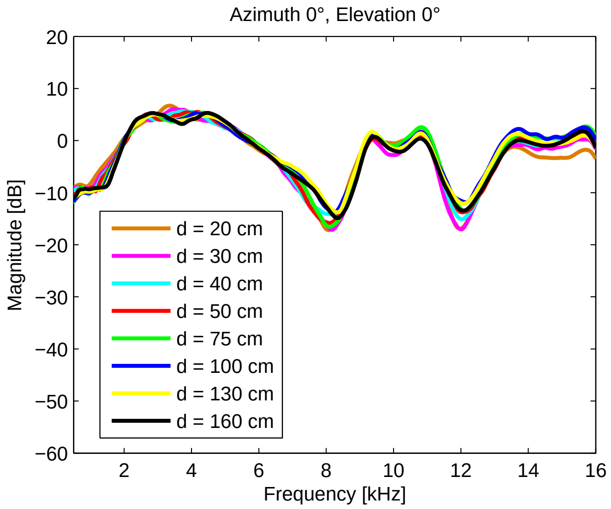

Figure 1. Normalized HRTFs for $\theta = 0°$ and $\phi = 0°$.

Spectral distortion $SD(\hat{H}(\theta_k, \phi_k, d), \hat{H}(\theta_k, \phi_k, d_8))$ can now be computed for each available angular coordinate[3] and each distance. The related results are represented in Figure 2, where the missing angular coordinates are conventionally assigned a 0-dB SD in the corresponding matrices' entries and coordinate $(\theta, \phi) = (0°, 90°)$, showing a SD never greater than 2.4 dB, is omitted for plotting issues.

A first look at these results reveals that the initial hypothesis is verified for the vast majority of the spatial coordinates: 71% of the nonzero entries are less than 3 dB and 87% are less than 4 dB, two SD values that reflect a reasonable agreement between different HRTFs considering both the inter-measurement variability and the increasingly lowpass behaviour of the human head as the sound source approaches that can be noticed back in Figure 1 for the highest frequencies. The latter effect is thought to be responsible for the average increase in SD for decreasing distances clearly detectable in Figure 2. Also notice how the responses on the median plane ($0°$-azimuth column) are always scarcely affected by distance: SD is never more than 4 dB.

Nevertheless, two major critical areas are shared by all of the seven plots:

1. a wide area (A_1) extending across several azimuth angles in the ipsilateral side of the head ($\theta = [0°, 180°]$) and concentrated around the horizontal plane, with a prominent tail around the coordinate $(130°, 40°)$; and

2. a more compact area (A_2) concentrated around the contralateral ear ($\theta = 270°$) at all elevations between $-40°$ and $40°$.

Here SD increases up to 9 dB for the closest distances and could imply the involvement of an effect of distance on the spectral features of the HRTF. Thus, further investigation is needed in order to understand the cause of such a systematic SD rise.

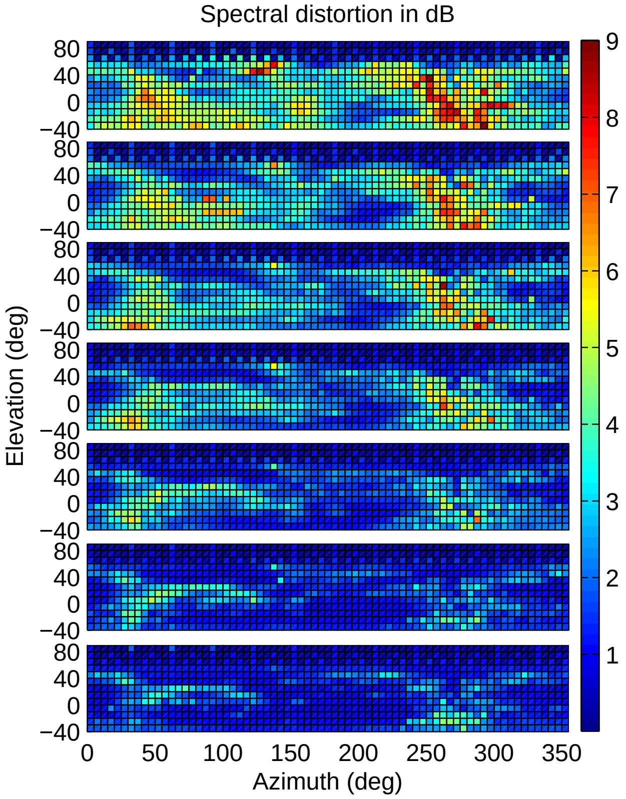

Figure 2. Spectral distortion between HRTFs at distance $d = 160$ cm and distance (top-to-bottom): $d = 20$ cm, $d = 30$ cm, $d = 40$ cm, $d = 50$ cm, $d = 75$ cm, $d = 100$ cm, $d = 130$ cm.

2.2 Deviation of spectral notches

We first examine what happens at two of the most critical angular coordinates, $(\theta, \phi) = (40°, 0°)$ in A_1 and $(\theta, \phi) = (265°, 0°)$ in A_2. The corresponding normalized HRTFs are traced in Figure 3 and Figure 4, respectively. What can be immediately seen in both figures is that the greatest dissimilarities among HRTFs are caused by the difference in both magnitude and frequency among the spectral notches and, to a minor extent, among the spectral peaks (as is the case of the peak around 9 kHz in Figure 4). Furthermore, some of the notches belong just to a strict subset of the 8 HRTFs, see e.g. the spectral notch at $14 - 15$ kHz appearing only for the farthest distances in Figure 3.

It should be then pointed out that, although the gross characteristics of the HRTF are preserved across distances, the presence/absence or displacement of some of the most important spectral cues for elevation detection could have an impact on localization. As a matter of fact, one could verify that notch shifts in the range of 1 kHz such as those appearing in Figure 4 usually correspond to an increase or decrease of the elevation angle greater than $20°$ [25]. Fur-

[3] Taking the vertical polar coordinate system as reference, elevation goes from $-40°$ to $90°$ in $10°$ steps, while azimuth goes from $0°$ to $355°$ in $5°$ steps except for elevation $60°$ ($10°$ steps), $70°$ ($15°$ steps), $80°$ ($30°$ steps), and $90°$ ($\theta = 0°$ only). The $(0°, 0°)$ direction is right in front of the listener, $(90°, 0°)$ is at the right ear, and $(270°, 0°)$ is at the left ear.

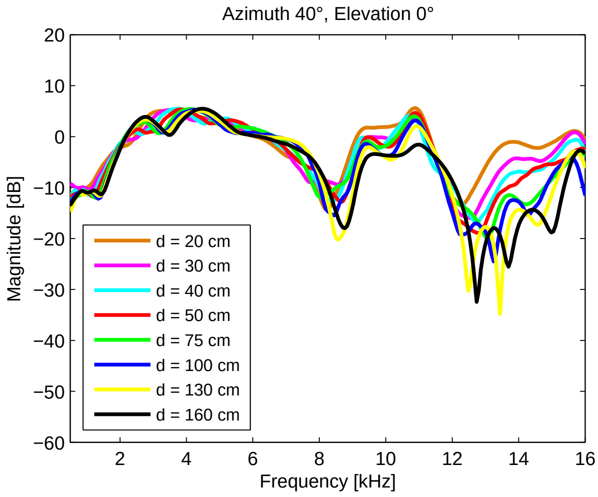

Figure 3. Normalized HRTFs for $\theta = 40^\circ$ and $\phi = 0^\circ$.

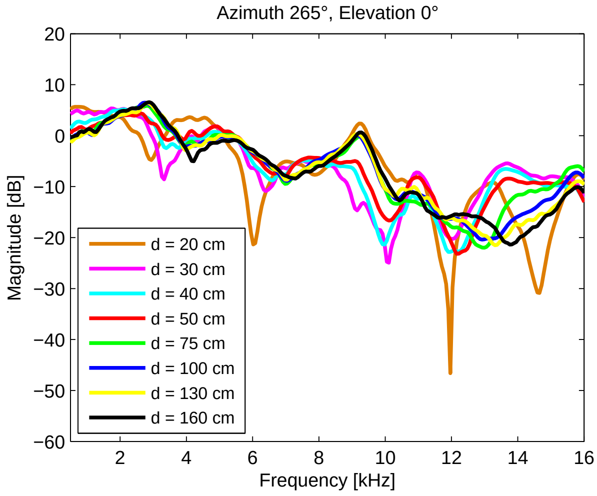

Figure 4. Normalized HRTFs for $\theta = 265^\circ$ and $\phi = 0^\circ$.

thermore, since the work of Moore *et al.* [11] we know that two steady notches in the high-frequency range differing just in center frequency are distinguishable on average if the mismatch is more than approximately 9% of the center frequency f_c of the lowest, regardless of notch bandwidth. [4] Although these results were found for just one frequency band (around 8 kHz), I may extend the validity of the assumption to the range $R_2 = [6, 11]$ kHz where the first two spectral notches, generally the most relevant in elevation perception [9], usually lie.

Using such assumption, another error measure is now introduced in order to attest whether HRTFs at different distances can potentially be distinguishable due to a frequency shift of one or more of their frequency notches. Having fixed the deepest notch in the range R_2 appearing in $H(\theta_k, \phi_k, d_8)$ as reference, let me define as the *notch frequency deviation* among the corresponding notches in the set of HRTFs $H(\theta_k, \phi_k, d_1), \ldots, H(\theta_k, \phi_k, d_8)$, where $d_1, \ldots, d_8$ are the eight available distances in increasing

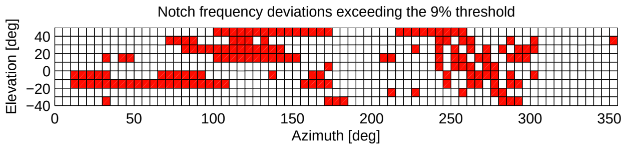

Figure 5. Notch frequency deviations exceeding the 9% threshold between $\phi = -40^\circ$ and $\phi = 40^\circ$.

order, and denote it as $dev(\theta_k, \phi_k)$, the difference in Hz between the frequency of the highest and the frequency of the lowest notch in the set, where available – if not (i.e. the notch is missing in one or more HRTFs), $dev(\theta_k, \phi_k)$ will conventionally take infinite value.

Each notch frequency deviation shall now be related to the aforementioned 9% threshold by simply expressing it as the percentual amount of the lowest notch frequency in the set. Figure 5 reports in a simple two-value matrix all those deviations that exceed the threshold, indicating a potential significant difference in the relative HRTF set, as positive (red) entries. In the figure, elevations greater than $\phi = 40^\circ$ are omitted because of the known lack of deep spectral notches for directions above the listener [25], where the notch frequency deviation metric loses its consistency.

Comparing this last matrix with those reported back in Figure 2 we can immediately notice a good agreement between the leftmost positive entries of the matrix and A_1, and a very good correspondence between the rightmost positive entries and A_2. The only significant differences between the two representations are

1. a greater number of positive entries around the tail of A_1, indicating that the notch frequency deviation metric begins to become inappropriate at medium-to-high elevations because notches are too shallow to be significant; and

2. a small number of positive entries for $\phi = -40^\circ$ and $\phi = -30^\circ$ meaning that, even though significant notch frequency deviations are rare in this range, the very deep notches appearing at low elevations greatly affect the SD computation even in presence of small notch frequency shifts.

Apart from these specific considerations, results suggest that the frequency deviation of spectral notches across distances is indeed the greatest source of spectral distortion among iso-directional HRTFs. Hence, the initial hypothesis on the independence between spectral elevation cues and distance is in this case not guaranteed for directions included in the previously defined areas A_1 and A_2. From the viewpoint of the listener, modifications of spectral features for sources close to the contralateral ear (area A_2) could be thought of having little effect on the perception of a sound source, as the ipsilateral ear will always receive a much louder signal from which it shall monaurally correctly decode the elevation of the source [2]. However, notch deviations in the ipsilateral side of the head (area A_1) could as well have an effect on elevation perception.

[4] By contrast, the perceptual relevance of changes in bandwidth and amplitude of a notch is little understood in previous literature.

3. CONCLUSIONS

By analyzing a recent database of distance-dependent head-related transfer functions, I found how the rough independence of elevation cues from distance advanced in [19] can not be attested at a purely analytical level for all directions of the sound source. The analysis was conducted on a single subject, a KEMAR manikin: KEMAR-related measurements (a smaller number of which were also used in [19] to support the incriminated claim) have the considerable advantage of being fully controllable, whereas similar measurements on a human subject would intolerably multiplicate the measurement time required to collect a standard set of HRTFs by the number of source distances, thus proportionally increasing the subject's tiredness and uneasiness. Nevertheless, a future contingent availability of distance-dependent HRTF sets measured on human subjects will allow a similar data analysis to be repeated, and the presented results to be further verified.

4. ACKNOWLEDGMENTS

I would like to acknowledge Michele Geronazzo and Dr. Federico Avanzini for the precious conversations throughout these last years that allowed the present work to be conceived and realized.

5. REFERENCES

[1] M. B. Gardner and R. S. Gardner, "Problem of localization in the median plane: Effect of pinnae cavity occlusion," *J. Acoust. Soc. Am.*, vol. 53, no. 2, pp. 400–408, 1973.

[2] M. Morimoto, "The contribution of two ears to the perception of vertical angle in sagittal planes," *J. Acoust. Soc. Am.*, vol. 109, pp. 1596–1603, April 2001.

[3] J. Hebrank and D. Wright, "Are two ears necessary for localization of sound sources on the median plane?," *J. Acoust. Soc. Am.*, vol. 56, pp. 935–938, September 1974.

[4] E. A. G. Shaw and R. Teranishi, "Sound pressure generated in an external-ear replica and real human ears by a nearby point source," *J. Acoust. Soc. Am.*, vol. 44, no. 1, pp. 240–249, 1968.

[5] E. A. Lopez-Poveda and R. Meddis, "A physical model of sound diffraction and reflections in the human concha," *J. Acoust. Soc. Am.*, vol. 100, pp. 3248–3259, November 1996.

[6] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization*. Cambridge, MA, USA: MIT Press, 1983.

[7] D. W. Batteau, "The role of the pinna in human localization," *Proc. R. Soc. London. Series B, Biological Sciences*, vol. 168, pp. 158–180, August 1967.

[8] D. Wright, J. H. Hebrank, and B. Wilson, "Pinna reflections as cues for localization," *J. Acoust. Soc. Am.*, vol. 56, pp. 957–962, September 1974.

[9] K. Iida, M. Itoh, A. Itagaki, and M. Morimoto, "Median plane localization using a parametric model of the head-related transfer function based on spectral cues," *Appl. Acoust.*, vol. 68, pp. 835–850, 2007.

[10] E. A. G. Shaw, "Acoustical features of human ear," in *Binaural and Spatial Hearing in Real and Virtual Environments*, pp. 25–47, Mahwah, NJ, USA: R. H. Gilkey and T. R. Anderson, Lawrence Erlbaum Associates, 1997.

[11] B. C. J. Moore, S. R. Oldfield, and G. J. Dooley, "Detection and discrimination of spectral peaks and notches at 1 and 8 kHz," *J. Acoust. Soc. Am.*, vol. 85, pp. 820–836, February 1989.

[12] J. Hebrank and D. Wright, "Spectral cues used in the localization of sound sources on the median plane," *J. Acoust. Soc. Am.*, vol. 56, pp. 1829–1834, December 1974.

[13] A. Wilska, *Studies on Directional Hearing*. English translation, Aalto University School of Science and Technology, Department of Signal Processing and Acoustics, 2010. PhD thesis originally published in German as "Untersuchungen über das Richtungshören", University of Helsinki, 1938.

[14] F. Asano, Y. Suzuki, and T. Sone, "Role of spectral cues in median plane localization," *J. Acoust. Soc. Am.*, vol. 88, pp. 159–168, July 1990.

[15] P. Zahorik, D. S. Brungart, and A. W. Bronkhorst, "Auditory distance perception in humans: a summary of past and present research," *Acta Acustica united with Acustica*, vol. 91, pp. 409–420, 2005.

[16] D. R. Begault, *3-D Sound for Virtual Reality and Multimedia*. San Diego, CA, USA: Academic Press Professional, Inc., 1994.

[17] M. B. Gardner, "Distance estimation of 0° or apparent 0°-oriented speech signals in anechoic space," *J. Acoust. Soc. Am.*, vol. 45, no. 1, pp. 47–53, 1969.

[18] D. H. Mershon and J. N. Bowers, "Absolute and relative cues for the auditory perception of egocentric distance," *Perception*, vol. 8, no. 3, pp. 311–322, 1979.

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

[20] T. Qu, Z. Xiao, M. Gong, Y. Huang, X. Li, and X. Wu, "Distance-dependent head-related transfer functions measured with high spatial resolution using a spark gap," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 17, pp. 1124–1132, August 2009.

[21] V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano, "The CIPIC HRTF database," in *Proc. IEEE Work. Appl. Signal Process., Audio, Acoust.*, (New Paltz, New York, USA), pp. 1–4, 2001.

[22] G. Eckel, “Immersive audio-augmented environments - the LISTEN project,” in *Proc. 5th IEEE Int. Conf. Info. Visualization (IV’01)*, (Los Alamitos, CA, USA), pp. 571–573, 2001.

[23] A. Kan, C. Jin, and A. van Schaik, “A psychophysical evaluation of near-field head-related transfer functions synthesized using a distance variation function,” *J. Acoust. Soc. Am.*, vol. 125, pp. 2233–2242, April 2009.

[24] M. Otani, T. Hirahara, and S. Ise, “Numerical study on source-distance dependency of head-related transfer functions,” *J. Acoust. Soc. Am.*, vol. 125, pp. 3253–3261, May 2009.

[25] S. Spagnol, M. Hiipakka, and V. Pulkki, “A single-azimuth pinna-related transfer function database,” in *Proc. 14th Int. Conf. Digital Audio Effects (DAFx-11)*, (Paris, France), September 2011.