Silke Bögelein, Fabian Brinkmann, David Ackermann, Stefan Weinzierl

Localization cues of a spherical head model

Bögelein, S.; Brinkmann, F.; Ackermann, D.; Weinzierl, S. (2018): Localization cues of a spherical head model. In: Fortschritte der Akustik - DAGA 2018 : 44. Jahrestagung für Akustik, 19.-22. März 2018 in München. Berlin: Deutsche Gesellschaft für Akustik e.V., 2018. pp. 347–350.

Localization cues of a spherical head model

Silke Bögelein, Fabian Brinkmann, David Ackermann, Stefan Weinzierl
silke.boegelein@campus.tu-berlin.de, {fabian.brinkmann, david.ackermann, stefan.weinzierl}@tu-berlin.de
Audio Communication Group, Technical University Berlin, D-10587 Berlin, Germany

Introduction

The human auditory organ, amongst many other physiologic functions, allows the listener to locate sound sources in three-dimensional space. This is either done with so called binaural cues, such as interaural level difference (ILD) and interaural time difference (ITD) or with monaural spectral cues (SC) [1]. It is widely accepted, that the accurate localization of sound source elevation relies on elevation dependent spectral cues that originate form the fine structure of the human outer ear. In addition, weaker elevation cues originate from reflections and shadowing at the human shoulder and torso, head-induced spectral cues, and from the ITD fine structure [2, 3, 4]. These localization cues are static, i.e. they exist without any motion of the source and listener. In contrast, motion cues stem from movements of the source or listener, and can be interpreted as motion induced temporal changes to the static localization cues. McAnally *et al.* [5] showed that left/right head-movements over an angular range of 16-32° considerably reduce the localization error and almost eliminate front-back confusion.

Localization cues are exploited by binaural technology to create natural sounding acoustic simulations that allow for accurate sound source localization. In most cases, these simulations are based on head-related transfer functions (HRTFs) that contain all cues mentioned above. Instead of using HRTFs, some applications rely on transfer functions obtained from a rigid sphere (spherical head transfer function - SHTF), in case HRTFs are not available, to avoid coloration or localization errors caused by non-individual HRTFs, for loudspeaker reproduction, or to render dynamic pseudo binaural auralizations (motion tracked binaural - MTB [6]). In these cases, the elevation localization is poor for static scenes, because of missing pinnae related spectral cues, but might be improved if motion cues come into play.

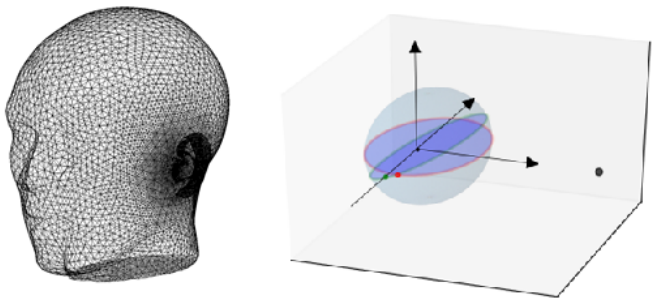

Figure 1: Left: Mesh of KEMAR's head; Right: sketch of the spherical off-center head model with offset-ears;

Up to now, the relevance and effect of motion cues has not been studied for the case of a spherical head model. The current study thus investigates the similarity of static and motion cues found in HRTFs and SHTFs, and introduces a simple model for estimating the localization error in the median plane based on dynamic ITD and ILD cues.

Method

KEMAR HRTFs

In this study the head of a KEMAR mannequin was used to assess human localization abilities. Therefore, a mesh of KEMAR's head was created as described in [7] and a gliding grid size between 1 mm (ipsilateral ear) and 10 mm (contralateral side) was rendered via the open source tool gmsh [8, 9]. The non-uniform discretization was chosen due to reasons of computing capacity. The left side of Figure 1 shows the mesh of KEMAR's head. Head related transfer functions were simulated numerically using the open source project Mesh2HRTF [10]. The tool allows to read the rendered geometrical mesh data, calculates the sound-field via the Boundary Element Method (BEM) and outputs the corresponding HRTF in the SOFA format. In our case a 3-dimensional Burton-Miller collocation BEM (reverberant Neumann boundary conditions) to avoid irregular frequencies [10] with a 10 m source distance was used. HRTFs were calculated reciprocally in a frequency range from 100 Hz to 22 kHz with a step size of 100 Hz. A Lebedev-Laikov-Grid of 35th order was chosen, due to nearly equally spaced

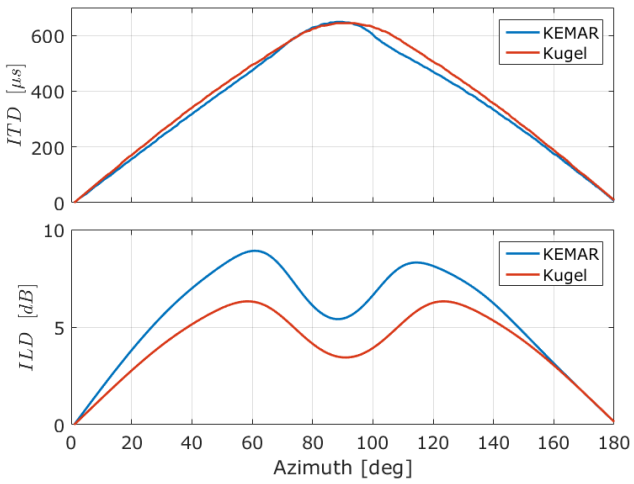

Figure 2: ITD (top) and ILD (bottom) of KEMAR and the spherical head model in the horizontal plane ($\vartheta = 0°$).

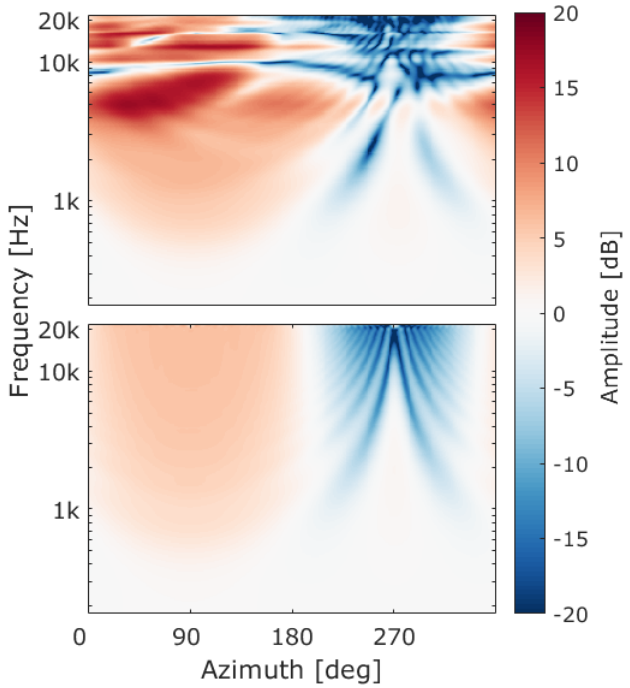

Figure 3: HRTFs (top) and SHTFs (bottom) of KEMAR and the spherical head model in horizontal plane ($\vartheta = 0°$).

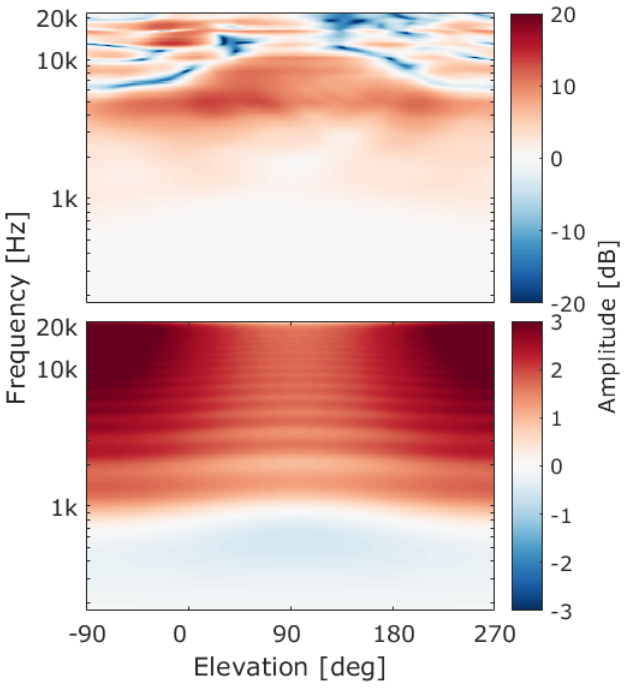

Figure 4: HRTFs (top) and SHTFs (bottom) of KEMAR and the spherical head model in horizontal plane ($\varphi = 0°$).

nodes on a sphere [11]. As it was taken for granted that KEMAR's head is axial symmetrical, HRTFs were simulated for the left ear only.

Spherical head model SHTFs

Ziegelwanger's time of arrival (TOA) model [12] was used to obtain the dimensions of the sphere. It estimates the radius and center of a sphere, as well as the azimuth and elevation angles of the ears by minimizing the squared error between TOAs estimated from the KEMAR HRTFs and the corresponding spherical TOA model, based on a non-linear least-squares solver. This resulted in an ear position of $\varphi_e = 89.62°$ azimuth and $\vartheta_e = -4.87°$ elevation, and sphere with a radius of $a = 8.35\,\text{cm}$. Then, considering the interaural centre to be the centre of the coordinate system [13], the spherical head was moved by $\sin(\varphi_e - \pi/2) \cdot a$ in x-direction and by $\sin(\vartheta_e) \cdot a$ in z-direction.

SHTFs were calculated analytically based on the spherical head model of Duda and Martens [14]. The existing model was expanded for off-center calculation and offset ear-positions. The analytical solution is given by

$$H(\varrho_0, \varrho, \mu, \Theta) = -\frac{\varrho_0}{\mu} e^{-j\mu\varrho_0} \sum_{m=0}^{\infty} (2m+1) P_m(cos(\Theta)) \frac{h_m(\mu\varrho)}{h'_m(\mu)}$$

with

$$\mu = \frac{\omega}{c} a = ka\,, \qquad \varrho = \frac{r}{a}\,, \qquad \varrho_0 = \frac{r_0}{a}$$

where Θ is the angle of incidence between the source and the ear position, $\omega = 2\pi f$ (f [Hz]: frequency), c [m/s] the speed of sound, a [m] the sphere's radius, r [m] the distance of the source to the centre of the (off-center) sphere, and r_0 [m] the distance of the origin of coordinates (in this case 10 m). With the off-center model, the two radii r and r_0 are distinguished. The incidence angle Θ between the ear and source position was calculated using the great circle distance

$$\Theta = \arccos(\sin(\vartheta)\sin(\vartheta_e) + \cos(\vartheta_e)\cos(\vartheta)\cos(\varphi - \varphi_e)).$$

The right side of Figure 1 shows a sketch of the extended spherical off-center head model with offset-ears. For further information see `AKsphericalHead.pdf` inside AKtools [15].

Static cues

Binaural cues

Binaural cues result in differences between both ears. They are represented by interaural time differences (ITD) and interaural level differences (ILD). The upper graph of Figure 2 shows the ITD, the lower graph the ILD in the horizontal plane for a sound source with 0° elevation. The broadband ITD was estimated from the low-passed ($f_c = 3$ kHz) and ten times up-sampled time signals using a threshold based onset detection (threshold -20 dB) [16]; the broadband ILD was estimated from the logarithmic RMS level differences between the left and right ear. Comparing the ITDs of KEMAR and the spherical head, only small differences can be seen throughout horizontal plane. This can be explained by matching the sphere's geometry to KEMAR's TOAs. However, the ILD of the sphere underestimates that of the KEMAR, as the influence of the pinnae on the ILDs in the case of the sphere is not present. A more detailed analysis of static cues can be found in [12, 17].

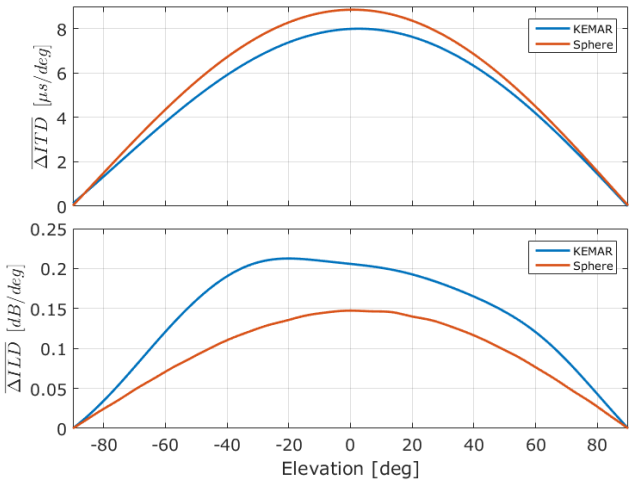

Figure 5: Change in ITD (top) and ILD (bottom) per degree for sources in the median plane, and head/source movement in the range of $|\varphi| \leq 10°$.

Spectral cues

Spectral cues are most important for localization in sagittal planes. Nevertheless, they also give information about left/right localization in transverse planes. Figure 3 shows HRTFs of KEMAR (upper graph) and SHTFs of the sphere (lower graph) in the horizontal plane. Compared to KEMAR, the fine structure is absent at high frequencies due to the pinnae absence. Additionally, there is a slight loss in the maximal amplitude for the ipsilateral side of the SHTFs. In general, however, HRTFs and SHTFs agree well in horizontal plane. Both show higher amplitudes at the ipsilateral side, and lower amplitudes at the contralateral side. Moreover, the inverse v-shape amplitude pattern at contralateral angles, due to shading of the (spherical) head at high frequencies, show up in both cases. Thus, left/right localization should be possible with spectral cues from a spherical head.
As already mentioned, the influence of the pinnae on the localization ability in median plane is said to be crucial for human hearing. The upper graph in Figure 4 shows HRTFs of KEMAR's head in median plane. The lower graph shows SHTFs in median plane. Note, that the SHTF amplitude was restricted to ±2.5 dB. Due to the missing pinnae, spectral features in median plane of the sphere's SHTF with offset-ears are a lot weaker than the well known spectral cues of KEMAR's head.

However, weak spectral cues in the SHTFs are the result of shifting the ears out of the principal axis (from $\varphi_e = 90°$, and $\vartheta_e = 0°$ to $\varphi_e = 89.62°$, and $\vartheta_e = -4.87°$), as it creates an asymmetry within an all-axis-symmetric geometry. Those weak spectral features, are in a range of 3 dB at high frequencies and might aid the elevation localization. However, these cues are almost identical for sources above and below the horizontal plane.

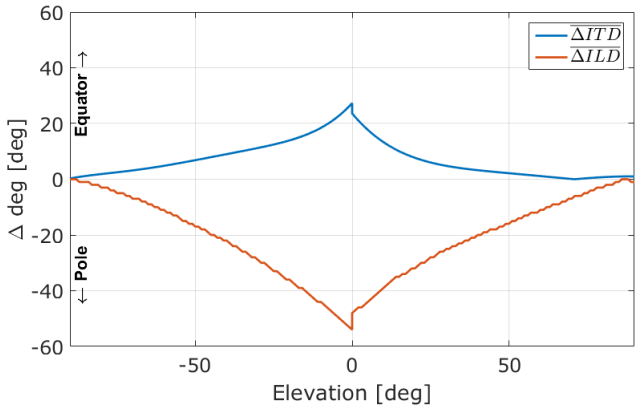

Figure 6: Estimated localization error caused by dynamic ITD and ILD cues of the spherical head model.

Motion cues

Motion cues are localization cues induced through head or source movements that result in changes of the ITD, ILD, and SC. In this study, we focus on sources in the sagittal median plane, and dynamic ITD and ILD cues that are caused by movements to the left and right, i.e. movements in transverse planes. They were analyzed by means of the ILD/ITD change per degree head/source movement averaged in the range of $|\varphi| \leq 10°$. ITD and ILD changes for sources in the median plane are shown in Figure 5.

In general, the dynamic cues are largest for sources on the horizontal plane, and decrease with increasing distance to the horizontal plane. However, opposite trends can be observed for ITD and ILD: Because the sphere's diameter overestimates KEMAR's head width – but underestimates KEMAR's head depth – the dynamic ITD cues of the spherical head are larger than those of KEMAR. In contrast, the dynamic ILD cues are smaller for the spherical head, and do not follow the asymmetry observed in the KEMAR data, which is caused by the missing pinnae.

It is likely to assume that the human auditory system learned these dynamic localization cues, and that a deviation from the learned features results in a localization error. For example, KEMAR expects a ΔITD of 6 µs/° at -39° elevation, but the spherical head delivers 6 µs/° at -47°, which results in a localization error of 8°. Figure 6 shows the estimated localization error in the sagittal median plane. Motion cues related to ITD result in a localization shift towards the horizontal plane of up to 30°, whereas ILD related motion cues cause errors of up to 50° that cause a shift towards the poles, i.e., away from the horizontal plane.

Discussion and conclusion

We assessed static and dynamic localization cues of a spherical head model with offset-ears with respect to their contribution to source localization in comparison

to HRTFs of a KEMAR dummy head. While static cues of the spherical head model provide sufficient information about the left/right position, the included localization cues in the median plane are rather weak. Fortunately, movement induced dynamic cues are a promising candidate for providing the missing information. In the case of left/right localization, the static ILD cues are underestimated by the spherical head model, but it might be assumed that the more realistic ITD cues dominate the perceived source location in this case [18]. The dynamic cues for the up/down localization are conflicting, and further investigations are needed to find out whether the conflicting cues balance each other, or if the auditory system prefers one cue over the other.
Future work will extend the analyses of this study to source positions outside the median plane, and to different geometries including a head without ears, an ellipsoid head model, and an isolated ear. Moreover, localization experiments will be carried out to separate the relevance and effect of static vs. dynamic localization cues and pinnae vs. head related cues. Furthermore, the influence of the reflection pattern of reverberant environments on the perceived source position could also be an interesting field of research in the context of spherical head models.

References

[1] Blauert, Jens (1997): *Spatial Hearing: The Psychophysics of Human Sound Localization.* Cambridge, MA, USA: MIT Press.

[2] Algazi, V. Ralph; Carlos Alvendano; and Richard O. Duda (2001): "Elevation localization and head-related transfer function analysis at low frequencies." In: *The Journal of the Acoustical Society of America*, **109**(3), pp. 1110–1122.

[3] Algazi, V. Ralph and Richard O. Duda (2002): "Approximating the head-related transfer function using simple geometric models of the head and torso." In: *The Journal of the Acoustical Society of America*, **112**(5), pp. 2053–2064.

[4] Benichoux, Victor; Marc Rébillat; and Romain Brette (2016): "On the variation of interaural time differences with frequency." In: *The Journal of the Acoustical Society of America*, **139**(4), pp. 1810–1821.

[5] McAnally, Ken I. and Russell L. Martin (2014): "Sound localization with head movement: implications for 3-d audio displays." In: *Frontiers in Neuroscience*, **8**, p. 210.

[6] Algazi, V. Ralph; Richard O. Duda; and Dennis M. Thompson (2004): "Motion-Tracked Binaural Sound." In: *J. Acoust. Soc. Am.*, **52**(11), pp. 1142–1156.

[7] Brinkmann, Fabian; et al. (2017): "A High Resolution and Full-Spherical Head-Related Transfer Function Database for Different Head-Above-Torso Orientations." In: *J. Audio Eng. Soc*, **65**(10), pp. 841–848.

[8] Geuzaine, Christophe and Jean-François Remacle (2009): "Gmsh: a three-dimensional finite element mesh generator with built-in pre- and post-processing facilities." In: *International Journal for Numerical Methods in Engineering*, **79**(11), pp. 1309–1331.

[9] Ziegelwanger, Harald; Wolfgang Kreuzer; and Piotr Majdak (2016): "A-priori mesh grading for the numerical calculation of the head-related transfer functions." In: *Appl. Acoust.*, **114**, pp. 99–110.

[10] Ziegelwanger, Harald; Wolfgang Kreuzer; and Piotr Majdak (2015): "MESH2HRTF: An open-source software package for the numerical calculation of head-related transfer functions." In: *Proceedings of the 22nd International Congress on Sound and Vibration, Florence, IT.*

[11] Bernschütz, Benjamin (2016): *Microphone arrays and sound field decomposition for dynamic binaural recording.* Ph.D. thesis, Technische Universität Berlin.

[12] Ziegelwanger, Wolfgang and Poitr Majdak (2014): "Modeling the direction-continuous time-of-arrival in head-related transfer functions." In: *The Journal of the Acoustical Society of America*, **135**(3), pp. 1278–1293.

[13] Møller, Henrik (1992): "Fundamentals of binaural technology." In: *Applied Acoustics*, **36**, pp. 171–218.

[14] Duda, Richard O. and William L. Martens (1998): "Range dependence of the response of a spherical head model." In: *The Journal of the Acoustical Society of America*, **104**(5), pp. 3048–3058.

[15] Brinkmann, Fabian and Weinzierl Stefan (2017): "AKtools - An open software toolbox for signal acquisition, processing, and inspection in acoustics." In: *142nd AES Convention, Berlin, Germany*, pp. e–Brief 309.

[16] Andreopoulou, Areti and Brian F. G. Katz (2017): "Identification of perceptually relevant methods of inter-aural time difference estimation." In: *The Journal of the Acoustical Society of America*, **142**(2), pp. 588–598.

[17] Bomhardt, Ramona and Janina Fels (2016): "Mismatch between Interaural Level Differences Derived from Human Heads and Spherical Models." In: *140th AES Convention, Convention Paper 9522.* Paris, France.

[18] Wightman, Frederic L. and Doris J. Kistler (1992): "The dominant role of low-frequency interaural time differences in sound localisation." In: *The Journal of the Acoustical Society of America*, **91**(3), pp. 1648–1661.