Head-Related Transfer Functions and Head Tracking

Benjamin Oksanen
Aalto University
Master's Programme CCIS / AAT

benjamin.oksanen@aalto.fi

Abstract

Head-related transfer functions (HRTF) can be utilized in binaural synthesis to reproduce spatial sound scenes with headphones. The spatial sound scene may not be correctly localized or externalized because the sound scene moves simultaneously with the listener's head movements. Head tracking system can be utilized to hold the sound scene static by compensating the head movements in the binaural synthesis. Results indicate that head tracking reduces the localization error and the rate of reversals. It improves the localization when using generic HRTFs for binaural synthesis. The externalization of the sound scene is improved when head tracking is used and the listener moves their head.

1 Introduction

In recent years, headphones have become the main way of listening to audio for many people because of their mobility and private listening experience. However, headphone listening differs from conventional stereo loudspeaker listening due to the lack of cross-talk between the channels and the effect of the listening room. The sound is often localized inside the head which is unnatural and can cause fatigue.

Spatial audio has become more popular with the rise of augmented and virtual reality applications. In most applications, the spatial sound scene should be reproduced with some type of headphone. Spatial sound scenes can be reproduced with headphones utilizing head-related transfer functions in the binaural synthesis. The head-related transfer functions are a set of transfer functions from positions in the space to the listener's ears.

Whether reproducing stereo or spatial audio, the sound scene follows the head movements to which the human hearing system is not accustomed to in natural sounds or loudspeaker listening. Thus, the reproduction is not optimal, in particular in virtual reality applications. Head tracking systems can be used to keep the sound scene stationary by taking into account the head movements in the binaural synthesis.

This paper is divided into three sections. First, the binaural technique is overviewed. Then, the head-related transfer functions are defined and the measurements, spatial resolution and their interpolation are discussed. Finally, head tracking systems are presented and the effects of head tracking in binaural synthesis are discussed.

2 Binaural Technique

The human hearing system is able to determine the direction and distance of sound sources using binaural cues which include coloration, interaural time difference (ITD) and interaural level difference (ILD) [18]. The ITD and ILD cues determine the localization in the horizontal plane, while the coloration, a monaural cue, contributes to the elevation of the source. Binaural cues are the result attenuations, reflections and delays of sound due to the geometries of the human head, torso and pinna.

Binaural technique refers to the recording technique where the two input signals to the human hearing system, the sound pressures at the ear drums, are recorded and reproduced. The sound pressures at the ear drums include all binaural cues the human hearing needs for localization of sound sources. Thus, the spatial sound scene can be reproduced exactly if the sound pressures at the ear drums are known.

3 Head-Related Transfer Functions

Transfer functions from a sound source, in a known position in a space to the eardrums, are called head-related transfer functions (HRTF). HRTFs include the binaural cues that enable localization in human hearing. HRTFs can be used to position sound sources in the virtual auditory space for synthesizing binaural signals. Even though HRTF refers, strictly speaking, only to the transfer function, the term HRTF is often used to refer to the time domain representation, head-related impulse response (HRIR), as well.

There are several ways of defining the head-related transfer functions. The traditional choice for the reference pressure is at the center point of the head while the listener is not present [18]. The other pressure of the transfer function can be measured in various positions. Most important positions are at the ear drum, at the entrance of an open ear canal and at the entrance of a closed ear canal. The head-related transfer functions, respective to these measurement positions, are also

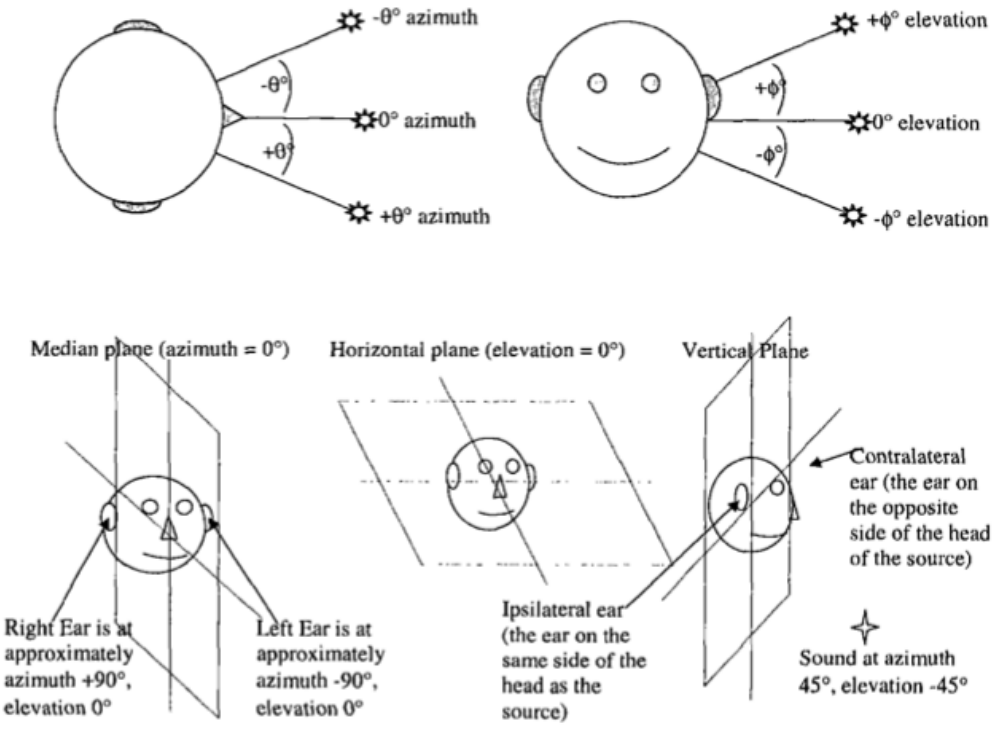

Figure 1: *Spatial coordinate system and terminology. Adapted from [4]*

known as free-field transfer functions [18]. HRTFs measured in the free-field are defined as

$$\text{HRTF}(\phi, \theta, r) = \frac{\text{sound pressure at the listener's ear}}{\text{sound pressure at the middle of the ear with listener absent}}$$

The HRTFs depend on the angles of incidence, elevation angle ϕ and azimuth angle θ, as well as the distance of the source, r. For far-field sources, when the pressure wave can be approximated as a plane wave, the distance to the source can be disregarded. Figure 1 shows the spatial coordinates in relation to the human head as well as some terminology related to HRTFs.

The same spatial information is present in all three measurement points and all of them yield the same results given that the binaural signals are reproduced with the correct ear canal effect [18]. This is due to the fact that the transmission path from the source to the ear drum is three-dimensional until the entrance of the ear canal after which the transmission becomes one-dimensional. The starting point of the direction-independent transmission is not clear but some studies suggest that the direction-independent region extends a few millimeters outside the entrance of the ear canal [8]. In the case of measuring outside a blocked ear canal, the transmission is three-dimensional until the blocked entrance and one-dimensional from the blocked entrance to the ear drum.

The effects of the ear canal resonance are highly individual due to the differences in the ear canal shape. Therefore, including the ear canal effect in the measurements, results in HRTFs that are individual to the subject and thus, may not work as well for other listeners. Measuring HRTFs at a position outside a blocked ear canal yields HRTFs that are most adaptable to other listeners as well as most practical to measure [19]. HRTFs with ear canal effects are also less practical to reproduce since the reproduction system must have a flat response at the ear drum position in order to avoid introducing the ear canal effects twice [18]. HRTFs measured outside a blocked ear canal are practical to reproduce with quality headphones that already approximate the ear canal effects correctly.

3.1 HRTF measurements

HRTFs measured from a single source position contain binaural cues only for that discrete position. For accurate reproduction of spatial sound scenes, HRTF data from all directions must available on a grid that corresponds to the angular resolution of the human hearing. The minimum audible angle, i.e. the smallest distinguishable change in direction of the sound, varies depending on the source location in relation to the head of the listener. The minimum audible angles vary between $1-10$ degrees depending on the direction and the frequency of the sound [15]. The well known cones of confusion, the cone shaped areas around azimuth angles $\theta = \pm 90^\circ$ where the direction cannot be resolved from ITD and ILD, have minimum audible angles of more than 40° [15] .

There are several methods for measuring HRTFs. Commonly used measurement setup consist of a rig which holds a loudspeaker that can be moved along a sphere with a radius of 1 to 2 meters around the stationary subject or dummy head with the two binaural microphones. The measurement is made with the source in one spatial position, after which the source is moved to the next position. Depending on the application, HRTFs can be measured in an anechoic chamber for free-field HRTFs or in a room for binaural room impulse responses (BRIR).

The number of measured spatial positions grows as the spatial resolution is increased. For example, measuring the HRTFs for elevation angles $\phi = -40... + 90^\circ$ in 10° steps and the azimuth angles $\theta = \pm 180^\circ$ in 5° steps results in 710 spatial positions [7]. Measuring with 5.6° angle step with a uniform sampling density on the sphere results in 1250 spatial positions [1]. The high number of spatial measurement points is problematic especially for individual HRTFs measured from a human subjects as the measurement is very time-consuming and the subject must remain still during the process.

HRTFs can be measured utilizing a constantly moving source instead of measuring at stationary positions. The measurement time is reduced significantly by using a moving source with angular speed of $2^\circ/s$ and $1s$ long sine sweep from 20 to 20 kHz

[21]. Even though the speaker moves 2° during each measurement, the resulting spatial blur is not audible.

One approach to measuring HRTFs utilizes a stationary loudspeaker in an anechoic room and instead of the loudspeaker, the human subject is rotated. The HRTFs can be estimated using a least-mean-square adaptive algorithm in combination with head-tracking system which provides the information of subject's head movements [9]. The head-tracking allows for unconstrained movements of the subject's head and the measurement approach still provides accuracy comparable to the measurements with continuously rotating loudspeaker [9].

Instead of measuring the binaural signals of a dummy head or a human subject, HRTFs can be acquired from a 3D model of the subjects head using for example finite difference method for simulating the acoustic wave propagation from different directions [17]. The 3D model can be obtained e.g. from magnetic resonance imaging data or from computer vision data.

3.2 Spatial Resolution of HRTFs

Given that HRTFs are measured as a function of elevation and azimuth angles at a certain distance, the spatial measurement positions are all on a surface of a sphere. Especially when measuring HRTFs of a human subject, the number of spatial measurement position must be limited in order to minimize the time required for the measurement. This can be achieved by choosing the most important positions and interpolation the remaining positions from the measured HRTFs [16].

Due to the shape of the sphere, distance on the sphere for a certain azimuth angle step is smaller in the poles than on the median plane. Therefore, using equal angle step for the measurement positions creates much more dense grid on the polar areas of the sphere which also happens to be the area where the minimum audible angles are larger [15]. There are ways to create more uniform grid of measurement positions on the surface of the sphere. One such sampling grid is the IGLOO presented in [26] that divides the polar areas of the sphere to three subregions each and the equatorial area to six subregions. Each subregion has 64 measurement positions. Figure 2 shows the measurement points of the IGLOO grid. Avoiding overly dense grid near the polar caps of the sphere is also beneficial as it reduces distortion when determining spherical harmonics from the HRTF data [26].

3.3 Interpolation of HRTFs

As it is time-consuming to measure HRTFs with high spatial resolution, interpolation or other numerical methods can be used to augment HRTFs measured with low spatial resolution. Efficient computation of HRTF at arbitrary points is especially crucial in real-time spatial audio and virtual reality applications.

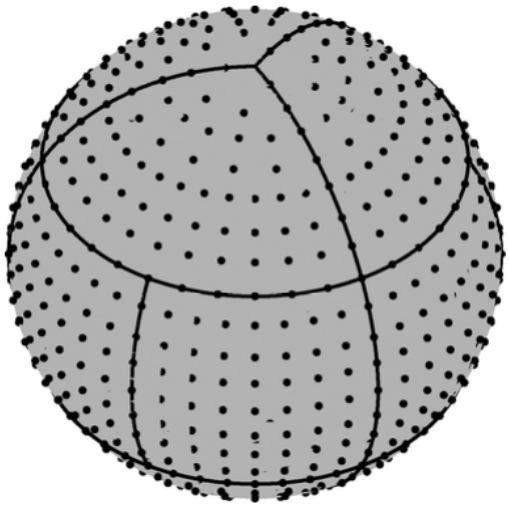

Figure 2: *IGLOO grid with 3:6:3 area division. Adapted from [26]*

Bilinear interpolation is a straightforward method of interpolation HRTFs. It computes the weighted average of four measured HRTFs surrounding the arbitrary point on the sphere to estimate the HRTF in that point [6]. The computation is done in time domain using the complementary HRIRs. Figure 3 shows the graphical interpretation of the bilinear transform where $\hat{h}$ is the HRIR at the arbitrary point, C_θ and C_ϕ are the relative angular positions of the arbitrary point and h_a, h_b, h_c and h_d are the measured HRIRs.

Numerous methods have been proposed to lighten the computational load by making the interpolation process more efficient. For example, the use of inter-positional transfer functions in the interpolation can make the interpolation process more efficient [6]. The inter-positional transfer functions are used two approximate two of the three HRTFs used in the interpolation process which reduces the computational load when low-order inter-positional transfer functions are used.

Another approach to interpolate the HRTF data is to create two spatial-variable dependent transfer function based on a continuous variable digital delay [5]. The method utilizes a polynomial interpolation to form a function for a subregion of the sphere where the angular coordinates correspond to the simulated HRTFs.

In [24], the authors propose using wavelet transform and triangular interpolation in wavelet domain. In [12], the authors reduce the HRTF data by applying principal component analysis and interpolate the HRTF using minimal state-space interpolation matrix. Deep learning and neural networks have been used together with different parametrical models to synthesize higher resolution HRTFs from sparse data, e.g. principal component analysis with fully-connected neural network [3].

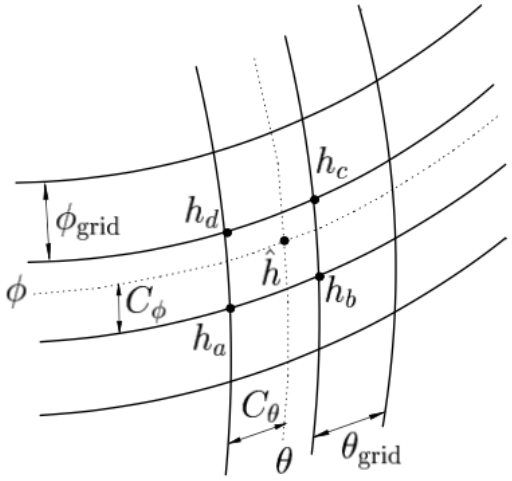

Figure 3: *Bilinear interpolation. Adapted from [6]*

4 Head Tracking in Binaural Synthesis

Binaural synthesis utilizing HRTFs can be used to reproduce e.g. a conventional stereo loudspeaker listening setup that is perceptually undistinguishable from the real loudspeaker listening setup. However, even small movements of the head can distort the illusion of the spatial sound scene because our hearing system does not expect the spatial sound scene to move along with the head movements. Head tracking must be utilized to keep the spatial sound scene stationary when the listener's head moves.

The human hearing system is capable locating sound sources accurately using only the binaural signals to the ears. Even though the localization is accurate, in some cases confusion occurs which are often resolved by small head movements.

4.1 Head Tracking Systems

Head tracking refers to tracking the movements and position of the human subject's head. The actual head tracking can be implemented e.g. with motion sensor attached to the head or with camera-based computer vision. Motion sensor head trackers are suitable for headphone reproduction since the head tracker can be attached to the headphone. The actual motion sensor can be implemented with acceleration sensors, magnetic field sensors or gyroscope sensors. All sensor types have their disadvantages, but the gyroscope sensor is the simplest and has least problems [20]. Camera-based head tracking systems are suitable for loudspeaker reproduction since the listener should be somewhat static in the constrained area between the speakers.

In spatial audio reproduction, head tracking is used to keep the spatial sound scene stationary when the listener moves their head. Head tracking for spatial audio reproduction using HRTFs should at least take into account the angular movements of the listener's head in the horizontal plane. Horizontal rotations of the head have the most substantial effect on the perceived spatial sound scene as the human hearing system is most accurate in the horizontal plane. Adding vertical head tracking allows the listener to tilt their head while keeping the spatial sound scene stationary although according to some studies vertical head head tracking does not offer significant benefits on the localization [14].

In virtual and augmented reality applications in particular, it can be meaningful to track also the position of the head in the space. This allows for the listener to move in the virtual sound scene. However, in most audio applications, the movement in space is not important or even desired, and therefore only the angular head movements are considered here.

The frequency at which the position of the head should be tracked for good quality reproduction depends on the individual. In [13] the authors found that some listeners could not recognize a difference between 4 Hz and 100 Hz update rates yet others could distinguish the difference between tested 18 Hz and 100 Hz update rates. Therefore, the update rate of head tracking should be at least higher than 18 Hz in order for it to be suitable for all listeners.

4.2 Reversals and Localization Error

Reversal refers to the front-back or back-front confusions that happen when the human hearing system cannot correctly decide from ITD and ILD cues of the sound source if the source is located in the front or back. ITD and ILD cues have a symmetry in the horizontal plane so that there a symmetric pairs of source location in front and back that have the same ITD and ILD. Small movement of the listener's head causes the ITD and ILD change according to their position which offers the human hearing system the information of the sound source to determine whether the source is located in front or in back. E.g. when rotating the head to the left, the left ear would move further away from a source in front and closer to a source in the back.

Moving sound sources do not reduce the reversal effect even though the ITD and ILD values change unless the movement of the source is controlled by the listener [25]. The knowledge of the direction of the movement offers the human hearing system the necessary information for determining the source location.

Along with reducing reversals, head movements reduce the localization error both in horizontal and vertical planes [23]. The up-down confusions are reduced by head movements as the monaural spectral cues change according to the head position.

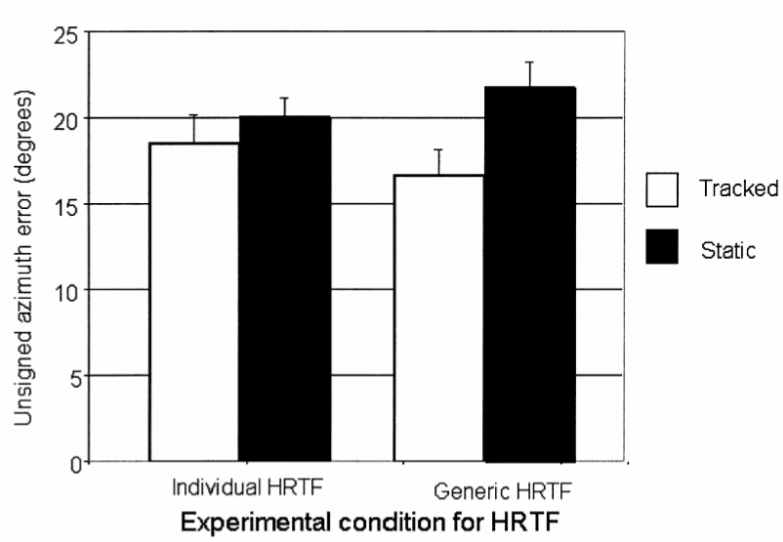

Figure 4: *Unsigned azimuth error mean values and standard error. Adapted from [2]*

The use of head tracking in HRTF binaural synthesis has been found to reduce the azimuth error in localization [2, 22]. Figure 4 shows the results of listening tests conducted in [2] for azimuth error in localization using both individually measured HRTFs and generic HRTFs. The reduction in azimuth errors using head tracking is more significant with generic HRTFs [2]. In [22], the authors found that the difference between generic and individual HRTFs was not significant but both cases improved with head tracking. The differences might arise from different HRTFs, listeners and stimuli. The head movements allow the human hearing system to calibrate to the generic HRTFs which is important since individual HRTFs are not easily available for everyone and many applications utilize generic HRTFs.

More importantly, the use of head tracking significantly reduced the rate of reversals in the horizontal plane [2]. The reversals are arguably more annoying for the listener than slight localization error. Nevertheless, head tracking has been proved to be beneficial for both reducing reversals and localization errors.

4.3 Externalization

Although in theory the spatial sound scene created with binaural synthesis using HRTFs should be externalized, i.e. the sound sources located at the distance and position where the HRTFs were measured, externalization does not always occur and the sound scene is localized inside the listener's head. The externalization may not happen or it can easily collapse if the listener's head moves and it is not accounted for in the binaural synthesis. Head tracking, which is used to keep the spatial sound scene stationary, improves the externalization given that some head movements are made [11].

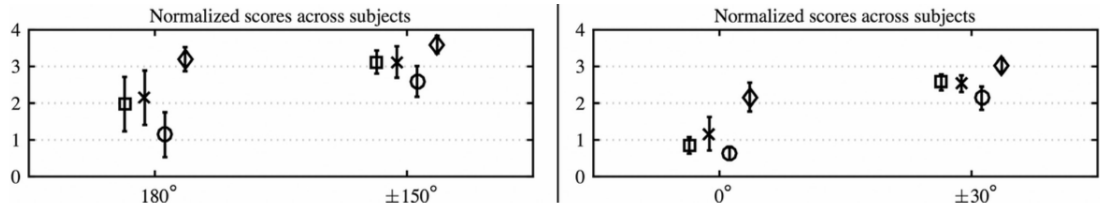

Figure 5: *Mean-normalized externalization scores with associated 95% confidence intervals. □: SØ (no head movement, no head tracking). ×: ST (no head movement, with head tracking). ○: MØ (with head movements, no head tracking). ◊: MT (with head movements, with head tracking). Adapted from [10]*

Figure 5 shows the mean externalization scores of multiple subjects for different azimuth angles from a listening tests carried out in [10]. The test consisted of four different conditions: no head movement without head tracking, no head movement with head tracking, head movement without head tracking, and head movement with head tracking. The externalization was graded on a six-point scale from 1 (The source is at the center of my head) to 5 (The source is externalized and remote).

The externalization of frontal and rear sources is considerably enhanced with head tracking when head movements are present in comparison to a situation where the listener's head is stationary. Furthermore, the externalization of the whole spatial sound scene is enhanced when applying head tracking to a listening situation where the listener's head moves. However, the head movements have to be sufficiently large, at least larger than the minimum audible angle, and voluntary to have an effect on the externalization [10].

5 Conclusions

Spatial audio reproduction with headphones relies on the use of head-related transfer functions which include the binaural and monaural localization cues of the human hearing. HRTFs are usually measured in an anechoic chamber from various positions on the surface of the sphere around the subject's head. The number and positions of the spatial measurement points affects the measurement time. The spatial resolution of the HRTF grid should be high enough for reproduction without introducing unwanted artifacts. HRTFs can be interpolated with various techniques from sparse data to reduce the measurement time while retaining high quality reproduction.

Head tracking systems can be used to keep the sound scene stationary while the head moves. The use of head tracking reduces localization errors and the rate of reversals. It also produces better externalization when compared to a static head without head tracking. Head tracking seems to improve the results whether using generic HRTFs or when using individual HRTFs although the improvement might be more significant with generic HRTFs according to some studies. Further studies into the matter are required.

References

[1] Algazi, V. R., Duda, R. O., Thompson, D. M., and Avendano, C. The CIPIC HRTF database. In *Proceedings of the 2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics (Cat. No.01TH8575)* (2001), pp. 99–102.

[2] Begault, D. R., Wenzel, E. M., and Anderson, M. R. Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source. *J. Audio Eng. Soc 49*, 10 (2001), 904–916.

[3] Bharitkar, S. G. Deep learning for synthesis of head-related transfer functions. In *Audio Engineering Society Convention 146* (Mar 2019).

[4] Cheng, C. I., and Wakefield, G. H. Introduction to head-related transfer functions (HRTFs): Representations of HRTFs in time, frequency, and space. In *Audio Engineering Society Convention 107* (Sep 1999).

[5] Freeland, F. P., Biscainho, L. W., and Diniz, P. S. HRTF interpolation through direct angular parameterization. In *2007 IEEE International Symposium on Circuits and Systems* (2007), IEEE, pp. 1823–1826.

[6] Freeland, F. P., Biscainho, L. W. P., and Diniz, P. S. R. Efficient HRTF interpolation in 3D moving sound. In *Audio Engineering Society Conference: 22nd International Conference: Virtual, Synthetic, and Entertainment Audio* (Jun 2002).

[7] Gardner, W. G., and Martin, K. D. HRTF measurements of a KEMAR. *The Journal of the Acoustical Society of America 97*, 6 (1995), 3907–3908.

[8] Hammershøi, D., and Møller, H. Sound transmission to and within the human ear canal. *The Journal of the Acoustical Society of America 100*, 1 (1996), 408–427.

[9] He, J., Ranjan, R., and Gan, W. Fast continuous HRTF acquisition with unconstrained movements of human subjects. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (March 2016), pp. 321–325.

[10] Hendrickx, E., Stitt, P., Messonnier, J.-C., Lyzwa, J.-M., Katz, B. F., and de Boishéraud, C. Influence of head tracking on the externalization of speech stimuli for non-individualized binaural synthesis. *The Journal of the Acoustical Society of America 141*, 3 (2017), 2011–2023.

[11] Inanaga, K., Yamada, Y., and Koizumi, H. Headphone system with out-of-head localization applying dynamic HRTF (head-related transfer function). In *Audio Engineering Society Convention 98* (Feb 1995).

[12] Keyrouz, F., and Diepold, K. A new HRTF interpolation approach for fast synthesis of dynamic environmental interaction. *J. Audio Eng. Soc 56*, 1/2 (2008), 28–35.

[13] Laitinen, M.-V., Pihlajamäki, T., Lösler, S., and Pulkki, V. Influence of resolution of head tracking in synthesis of binaural audio. In *Audio Engineering Society Convention 132* (Apr 2012).

[14] Mackensen, P., Fruhmann, M., Thanner, M., Theile, G., Horbach, U., and Karamustafaoglu, A. Head tracker-based auralization systems: Additional consideration of vertical head movements. In *Audio Engineering Society Convention 108* (Feb 2000).

[15] Mills, A. W. On the minimum audible angle. *The Journal of the Acoustical Society of America 30*, 4 (1958), 237–246.

[16] Minnaar, P., Plogsties, J., and Christensen, F. Directional resolution of head-related transfer functions required in binaural synthesis. *J. Audio Eng. Soc 53*, 10 (2005), 919–929.

[17] Mokhtari, P., Takemoto, H., Nishimura, R., and Kato, H. Comparison of simulated and measured HRTFs: FDTD simulation using MRI head data. In *Audio Engineering Society Convention 123* (Oct 2007).

[18] Møller, H. Fundamentals of binaural technology. *Applied acoustics 36*, 3-4 (1992), 171–218.

[19] Møller, H., Sørensen, M. F., Hammershøi, D., and Jensen, C. B. Head-related transfer functions of human subjects. *J. Audio Eng. Soc 43*, 5 (1995), 300–321.

[20] Pörschmann, C. 3-d audio in mobile communication devices: Methods for mobile head-tracking. *JVRB - Journal of Virtual Reality and Broadcasting 4*, 13 (2007).

[21] Pulkki, V., Laitinen, M.-V., and Sivonen, V. HRTF measurements with a continuously moving loudspeaker and swept sines. In *Audio Engineering Society Convention 128* (May 2010).

[22] Su, H., Marui, A., and Kamekawa, T. The effect of HRTF individualization and head-tracking on localization and source width perception in VR. In *Audio Engineering Society Convention 146* (Mar 2019).

[23] Thurlow, W. R., and Runge, P. S. Effect of induced head movements on localization of direction of sounds. *The Journal of the Acoustical Society of America 42*, 2 (1967), 480–488.

[24] Torres, J. C. B., and Petraglia, M. R. HRTF interpolation in the wavelet transform domain. In *2009 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics* (2009), pp. 293–296.

[25] Wightman, F. L., and Kistler, D. J. Resolution of front–back ambiguity in spatial hearing by listener and source movement. *The Journal of the Acoustical Society of America 105*, 5 (1999), 2841–2853.

[26] Zhang, W., Zhang, M., Kennedy, R. A., and Abhayapala, T. D. On high-resolution head-related transfer function measurements: An efficient sampling scheme. *IEEE Transactions on Audio, Speech, and Language Processing 20*, 2 (Feb 2012), 575–584.