BINAURAL RENDERING OF DYNAMIC HEAD AND SOUND SOURCE ORIENTATION USING HIGH-RESOLUTION HRTF AND RETARDED TIME

Christoph Urbanietz and Gerald Enzner

Institute of Communication Acoustics, Ruhr-Universität Bochum, 44801 Bochum, Germany
christoph.urbanietz@rub.de, gerald.enzner@rub.de

ABSTRACT

This paper is devoted to high-fidelity implementation of HRTF-based binaural rendering with fast head and source rotations in virtual acoustic reality. With an intuitive physical standpoint, we argue that head rotations should be rendered by a convolution model anchored in the sound receive-time. Conversely, the rendering of moving sound sources should be anchored in the sound emission-time, using retarded HRTF. Distant sound sources with bulk delay on the impulse response are handled efficiently by substituting the delay by additional retardation. The proposed rendering engine can utilize fast changing filter coefficients from a high resolution HRTF and satisfies some acoustic features as Doppler shift and dislocation inherently, which have been treated separately in conventional approaches or are neglected at all. To benchmark our rendering engine, we compare these acoustic features of the rendered signal with corresponding physical expectations.

Index Terms— Binaural Rendering, HRTF, Time-Varying

1. INTRODUCTION

Binaural sound rendering based on head-related transfer functions (HRTFs) has applications in hearing aids [1], voice communication, music performance, games [2], sonic detection and orientation [3], and virtual acoustic environments (VAE) using headphones [4–11]. The latter will benefit from head tracking [2] to unlock the sound-field from head rotation. For a realistic perception, the total system latency (TSL) has to be below 70 ms [12], although the actual threshold of the just noticeable TSL can be lower for the individual [13] or in the presence of a low latency reference signal [14].

For the rendering of a virtual binaural acoustic scene, HRTFs or their respective time-domain representation, i.e., the head-related impulse response (HRIR), are frequently used [4]. Many databases of HRTFs exist, but most of them exhibit a discrete spatial resolution [15–19]. There are also approaches that can deliver at least in the azimuth direction a quasi-continuous resolution [8, 20–22]. Nevertheless, in everyday practice mostly a discrete set of HRTFs is used for buffer-wise binaural rendering. In the case of head rotations or sound source motions this results in artifacts [23, 24], which can be partially overcome by switching strategies between HRTFs, e.g., simple switching, overlap-add or a fade-in-fade-out (cross-fading) method [7, 25]. Apart from these transient effects, additional factors such as insufficient reverberation, non-individualized HRTF, or head-tracking latency can affect the fidelity of perception.

In this paper, we pursue a physically-motivated idea of binaural rendering as far as it can be accomplished with measured HRTFs. In the light of increasing computational resource, our intention is to enhance the physical precision of the sound that would occur in the case of fast head movements or fast moving sound sources. The baseline of our approach is the time-domain linear convolution model of the sound propagation, but we will take other aspects of the time-varying nature of the acoustic system into account than this is usually done. As a benefit the proposed algorithm takes many physical effects into account inherently, which otherwise have to be handled separately in conventional binaural rendering techniques or are usually neglected. Particularly we will point out the contrast between head rotations and sound source rotations, i.e., movements around the head at constant distance. This contrast is not given attention in classical binaural rendering. The scope of this paper is the technique of the pure rendering and we rely on the availability of a continuous HRIR or an appropriate interpolation [26].

The remainder of the paper is organized as follows. Sec. 2 briefly revisits the state-of-the-art of buffer-wise and sample-wise binaural rendering. Sec. 3 then postulates improved processing according to the physics of head and sound source motion, respectively, while Sec. 4 experimentally confirms the proposed concepts.

This work is supported Nr. EFRE-0800372 NRW grant, LS-1-1-044d.

2. RELATION TO PRIOR ART

With a pure LTI-system we cannot create virtually moving sound sources or compensate head rotations in the context of HRTF-based binaural rendering, since the HRTF depends on the spatial orientation which changes with time and is therefore time-variant.

2.1. Buffer-wise rendering

Common practice, however, is to assume a piece-wise LTI-system on finite-length buffers of samples. For dynamic acoustic scenes these buffers employ different HRTF-filters and their outputs are cross-faded [7, 8, 25]. This can lead to artifacts, especially if the effectively cross-faded HRTF-filters differ massively from each other, e.g., when the head orientation changes fast, or if the spatial resolution of the HRTF is too coarse. Additionally a buffer-latency is added because samples have to be aggregated before being processed.

2.2. Sample-wise rendering

An approach to overcome some of these issues is to render the binaural sound sample-wise [27]. In this case the buffer-latency can be overcome. However, the more crucial point for us is that a sample-wise rendering can make use of sample-wise updates of the HRTF filters. This implies the assumption that an appropriate interpolation of discrete HRTF or a measurement of quasi-continuous HRTF data is available [21]. Furthermore, with the sample-wise update we can introduce a rendering scheme that considers the differences between moving sound sources and a moving head.

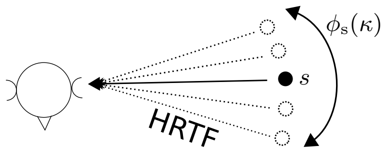


Fig. 1: Physical view of a rotating-source signal s, originating at angle ϕ_s at emission time κ, passing through HRTFs to a static head.

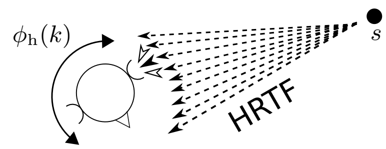


Fig. 2: Physical view of a static source s transmitting through the HRTF field while the rotating head selects the result at angle $\phi_h(k)$.

3. DYNAMIC BINAURAL RENDERING

We rely on the HRIR to describe the path of the signal from the source to the ear. This is a reasonable view, because the sound is traveling as a wave in the air with finite speed. We define $h(\ell, \Delta\phi)$ as the physically correct time-invariant HRIR at the relative angle $\Delta\phi = \phi_{\rm s(ource)} - \phi_{\rm h(ead)}$, with ℓ as coefficient index. First we will discuss the source motion and the head motion separately before combining them into a comprehensive convolution engine at the end.

3.1. Static head and moving sound source

First consider a model of a moving sound source successively emitting single samples that propagate along different paths to the ears. As the path is described by an HRIR-filter, each single source sample triggers a distinct HRIR, particularly the HRIR of the angle $\phi_{\rm s}(\kappa) - \phi_{\rm h}$ at emission time κ. This means that a single source sample $s(\kappa)$ originating at discrete time κ travels completely through the path described by $h(\ell, \phi_{\rm s}(\kappa) - \phi_{\rm h})$, where $\phi_{\rm h}$ is constant. At the head the result of all sound emissions is superimposed. The situation is illustrated in Fig. 1 and we describe the result at receiving-time k at one ear (the other in analogy) by

$$y_{\rm s}(k) = \sum_{\kappa=k-N+1}^{k} h(k-\kappa, \phi_{\rm s}(\kappa) - \phi_{\rm h}) s(\kappa) \tag{1}$$

$$= \sum_{\ell=0}^{N-1} h(\ell, \phi_{\rm s}(k-\ell) - \phi_{\rm h}) s(k-\ell) \tag{2}$$

where the relevant filter length of h is denoted by N and the second line is achieved by a substitution $k - \kappa \rightarrow \ell$. This form is equal to the algorithm proposed by [27]. We point out here that the angle of the HRIR being used is retarded by ℓ to the current time k.

3.2. Rotating head and static sound source

Now consider the signal of a static source traveling through all possible HRIR-filters simultaneously. Then, at the end of this path, the appropriate result is selected according to the current relative head position at receiving time k (not emission time κ as before). The situation is illustrated in Fig. 2. In this model, the output sample at time k only depends on the HRIR of the current angle $\phi_{\rm s} - \phi_{\rm h}(k)$. Hence, we can describe the signals at the ears by

$$y_{\rm h}(k) = \sum_{\ell=0}^{N-1} h(\ell, \phi_{\rm s} - \phi_{\rm h}(k)) s(k-\ell) \tag{3}$$

$$= \sum_{\kappa=k-N+1}^{k} h(k-\kappa, \phi_{\rm s} - \phi_{\rm h}(k)) s(\kappa). \tag{4}$$

The significant difference with respect to (1) and (2) is the use of the current angle index k instead of the retarded $\kappa = k - l$.

3.3. Simultaneous motion of head and sound source

With the arguments of both cases it is now straightforward to describe simultaneous rotations of head and sound source by

$$y_{\rm c(ombined)}(k) = \sum_{\kappa=k-N+1}^{k} h(k-\kappa, \phi_{\rm s}(\kappa) - \phi_{\rm h}(k)) s(\kappa) \tag{5}$$

$$= \sum_{\ell=0}^{N-1} h(\ell, \phi_{\rm s}(k-\ell) - \phi_{\rm h}(k)) s(k-\ell) \tag{6}$$

where (1) and (4) merged to (5), while (2) and (3) generalize to (6). These expressions collapse to (1-4) if either $\phi_{\rm s}$ or $\phi_{\rm h}$ were constant.

3.4. Adding distance between source and receiver

The significance of an appropriate distinction between head and sound source rotations is more emphasized with increasing distance r between head and source because then also the time retardation for the source angle and source signal is increasing. The distance r, and therefore the corresponding signal delay Δt, can simply be realized by padding $d = \Delta t \cdot f_{\rm s} = r/c \cdot f_{\rm s}$ leading zeros to the pure HRIR h, where $f_{\rm s}$ is the sampling frequency and c the speed of sound. As we restrict in this paper to angular motion only and therefore d does not change with time, we round d to integer. We call the new HRIR $h_{\rm d(elayed)}$. As there is no need to process zeros in the new HRIR, we can simply start the convolution (6) at $\ell = d$, or equivalently apply additional retardation d to the pure HRIR h:

$$y_{\rm c}(k) = \sum_{\ell=d}^{N-1+d} h_d(\ell, \phi_{\rm s}(k-\ell) - \phi_{\rm h}(k)) s(k-\ell) \tag{7}$$

$$= \sum_{\ell=0}^{N-1} h(\ell, \phi_{\rm s}(k-\ell-d) - \phi_{\rm h}(k)) s(k-\ell-d). \tag{8}$$

4. EXPERIMENTAL VERIFICATION

The proposed dynamic rendering algorithm assigns a distinct HRIR, corresponding to the relative angle, to each combination of input and output samples. The difference of head (4) and source rotations (1) is formally established by the presence of the receiving-time index k or the emission-time index κ with the angles $\phi_{\rm h}$ and $\phi_{\rm s}$, respectively. The presence of the emission time index in fact manifests the delayed reception due to the finite speed of sound. Under linear time-invariant conditions, i.e., with constant $\phi_{\rm h}$ and $\phi_{\rm s}$, the two cases will coincide. To evaluate the proposed algorithm we will thus render specific time varying scenes, corresponding to an input signal s, dynamic angles $\phi_{\rm h}$ and $\phi_{\rm s}$ and various fixed distances r. As shown in Fig. 3 the rendering output is then evaluated w.r.t. the physical expectation according to the input parameter space. For our calculations a continuous-azimuth HRIRs as a linear interpolation from a 1-degree sampled measurement as described in [21] is used.

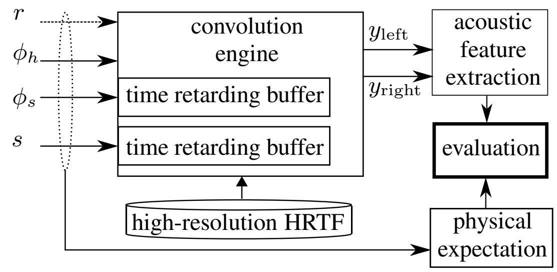


Fig. 3: Dynamic binaural rendering and verification.

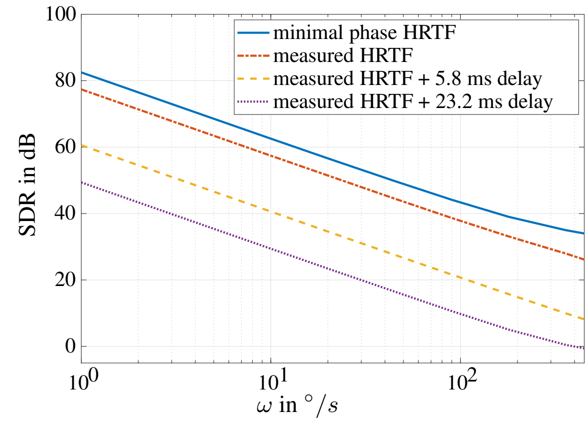


Fig. 4: SDR between head rotation and source rotation depending on rotation speed of the head for different filter bulk-delays.

4.1. Considering the Doppler shift

As a first way to assess the potential of our approach, the presence of the Doppler shift is investigated. As the ears are located out of the center of the head, both source and head rotations cause a time-varying source-to-receiver distance and, hence, a frequency shift of the source signal is expected. Since the Doppler effect is merely caused by the variation of distance and to form a physical expectation, we use a very simple head model taking only the interaural time difference (ITD) without shadowing effects into account [28]. The distance of a distant static sound source to the left/right ear is approximately $r \pm a\sin(\Delta\phi)$, where $a \ll r$ is the radius of the head, r the distance between the source and the center of the head, and $\Delta\phi$ the angle between source and head. For a rotation of head or source with a constant angular velocity ω we get as a maximum relative velocity of the ears towards and away from the source $v_{\max} = \max_t \frac{\partial}{\partial t}\left[r + a\sin(\omega t)\right] = a\omega$. In the case of a moving source this would result in a maximum and minimum frequency $f_{\text{s,min/max}} = \frac{f_0}{1 \pm v_{\max}/c}$ of the observed signal, whereas in the case of a rotating head we get $f_{\text{h,min/max}} = f_0\left(1 \mp v_{\max}/c\right)$ [29, 30]. Table 1 compares these expected frequency ranges with the measured frequency spectrum of the convolution output.

With a natural $a = 0.085\text{m}$, $c = 340\text{m/s}$, and an angular velocity $\omega = 2\pi/\text{s}$ the maximum velocity is $v_{\max} = 0.53\text{m/s}$. For a sinusoidal source signal with $f_0 = 5\text{kHz}$ the corresponding Doppler shifts are small as shown in Table 1a and below the human difference limen for pitch [31, 32]. The difference between moving head and moving source is even below the measurement resolution of 1 Hz. Therefore, to prove fundamental correctness in the reproduction of Doppler shifts, a theoretical 100 times bigger head is assumed as an academic example. With $a = 8.5\text{m}$ and $f_0 = 440\text{Hz}$ the Doppler shift becomes more significant as shown in Table 1b. In this case also the difference between head and source rotations are demonstrated and the rendering outputs give different frequency ranges for both cases as physically expected. Apart from possible errors within the resolution the expected frequency range is replicated by the rendering and therefore we can claim for our proposed algorithm that even the non-linear Doppler-effect is well contained.

		source rotation		head rotation	
		$\frac{f_{\text{s,min}}}{\text{Hz}}$	$\frac{f_{\text{s,max}}}{\text{Hz}}$	$\frac{f_{\text{h,min}}}{\text{Hz}}$	$\frac{f_{\text{h,max}}}{\text{Hz}}$
(a)	expected	4992.2	5007.9	4992.1	5007.9
	measured	4992	5008	4992	5008
(b)	expected	380.3	522.0	370.9	509.1
	measured	380	522	370	510

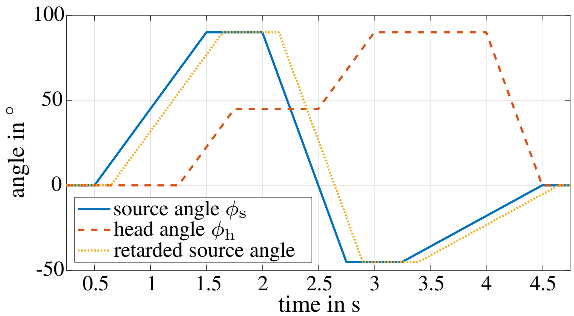
Table 1: Frequency range of the Doppler shift measured from rendering and as physical expectation. Case (a) with $a = 0.085\text{m}$ and $f_0 = 5\text{kHz}$, case (b) with $a = 8.5\text{m}$ and $f_0 = 440\text{Hz}$.

4.2. SDR comparison of head and source rotation

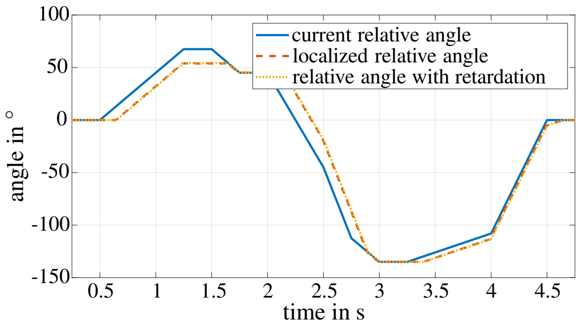
Although it is desirable to verify the rendered output waveform with a physical measurement this would be very difficult due to synchronization issues and noise. For an inner evaluation of the rendering, we can still compare the rendered output signal of a source rotating counter-clockwise and a head turning clockwise. The signals are compared by a broadband waveform distortion measure, i.e., signal-to-distortion ratio SDR $= \sqrt{\mathcal{E}\{y_{\text{s}}^2(k)\}\mathcal{E}\{y_{\text{h}}^2(k)\}}/\mathcal{E}\{(y_{\text{s}}(k) - y_{\text{h}}(k))^2\}$, where $y_{\text{s}}(k)$ and $y_{\text{h}}(k)$ are obtained from the moving-source (1) and from the rotating-head (4) implementation, respectively, and a full 360° revolution of a white-noise signal $s(k)$ is applied in both cases. Fig. 4 demonstrates considerable SDR contrast as a function of the angular velocity ω and source distance r. A measured HRTF [20] is employed with additional delays of Δt = 5.8 ms and Δt = 23.2 ms, corresponding to r = 2.0 m and r = 7.9 m additional distances. In a fourth case the measured HRTF is shifted in time to achieve minimum phase. It can be seen that for low ω the SDR rises to infinity, which confirms the equality of (1) and (4) for quasi time-invariance, while the fast movement in the order of $\omega = 450°$/s results in a waveform distortion in the order of 34 dB, 26 dB, 8 dB and -1 dB, respectively, which substantiates the difference of head and source rotations. Particularly we see that the SDR mainly depends on the product of $\omega\Delta t$, wherefore the SDR curves of the cases Δt = 5.8 ms and 23.2 ms are shifted by approximately a factor 4 on the ω-axis. The product $\omega\Delta t$ can be interpreted as an angular dislocation to be further investigated in Section 4.3.

4.3. Sound source localization

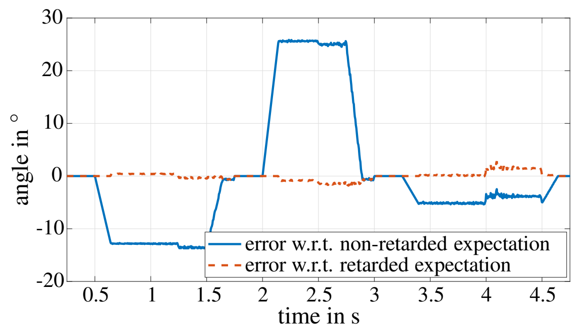
Although the former test based on the SDR confirms the fundamental difference between rotating head and source it may not be the right measure to quantify audible artifacts. As an example, a small time- or phase-shift can dramatically decrease the SDR measure, whereas the effect is not noticeable by the human sense of hearing.



(a) Source and head trajectory



(b) Localized and predicted relative angle



(c) Localization error with respect to physical expectation

Fig. 5: Example of perceptual dislocation of 50 m distant sound.

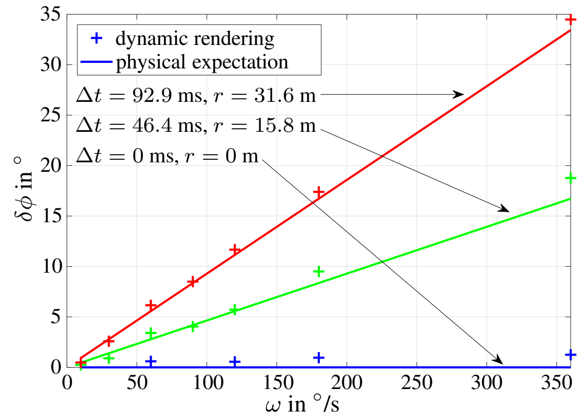


Fig. 6: Angle mismatch of localized source towards actual position.

A noticeable feature of a binaural sound, however, is the perceived angle of the sound source. In dynamic environments spatial sound perception includes the phenomenon of dislocation, i.e., the current source position will be perceived with a delay of $\Delta t = r/c$ due to the finite speed of sound. A commonly known example are planes. A plane flying in 10 km distance over the observer is heard at a position that the plane had approximately $\Delta t = 29.4$s before. For deeper illustration, consider a dynamic scene, where the head and a sound source rotate with various velocities at 50 m distance. Fig. 5a depicts the head angle, the source angle, and the source angle retarded by the time Δt the sound needs to travel. Fig. 5b shows the current relative angle (i.e. the difference between the current source angle and the current head angle), the relative angle between retarded source and current head orientation (i.e. our physical expectation) and an instrumental localization based on the rendered sound. We use a localizer in matched-filter form [33] based on the steered-response-power principle [34] in time-domain. It applies a full search on the available HRTF table with 0.1 degree resolution. The HRTF in each search direction is normalized to unit-norm [35]. Yet, the localizer will have an internal limitation of precision since the scene is dynamic and the localizer needs an appropriate number of samples (here: 1000) to localize the sound direction. For the configuration at hand we found at the highest angular velocity a mean accuracy of the localizer of 1.2 degree. Within these limits we see that the proposed algorithm renders sound that can be localized at the physically expected position as shown by the small location errors in Fig. 5c. Due to an appropriate handling of retarded time the rendering produces a sound coming from a physically correct position. Of course this retarded time can be taken separately into account when the sound is rendered by a classical method, but our algorithm inherently applies such treatment.

For a more global view on this sound source localization feature, Fig. 6 compares the physically expected dislocation for various angular velocities ω and various sound source distances r with that of our rendering/localization chain, averaged on a frontal 180° revolution of the sound source. The physical expectation of the dislocation is $\delta\phi_{\mathrm{phys.}} = \omega\Delta t = \omega r/c$. The measured dislocation of the dynamic rendered signal matches the expectation within the finite accuracy of the localizer. This shows that the proposed rendering algorithm can inherently handle this physical dislocation.

5. CONCLUSIONS

We introduced a binaural rendering algorithm for sound scenes with rotating head and sources. It can make use of high resolution HRTF to update the HRIR sample-wise, while head and source rotations are taken differently into account according to different physical mechanisms. The inner difference of the rendering between both rotations is evaluated in terms of SDR and it is shown that the difference increases with rotation speed and source distance. A correct reproduction of the Doppler-shift is shown and further substantiates the difference between head and source motion, even if the effect is small in usual cases. At the end we prove the inherently correct reproduction of the dislocation of moving sources, due to the finite travel time of sound from source to receiver. We thus come a step closer to a correct reproduction of binaural scenes without the need of taking many effect separately into account, since the proposed algorithm inherently implies them. Especially the difference between head and source motion is elaborated and confirmed with simulations.

6. REFERENCES

[1] S. Doclo, W. Kellermann, S. Makino, and S. E. Nordholm, "Multichannel signal enhancement algorithms for assisted listening devices: Exploiting spatial diversity using multiple microphones," *IEEE Signal Processing Mag.*, vol. 32, no. 2, pp. 18–30, March 2015.

[2] V. R. Algazi and R. O. Duda, "Headphone-based spatial sound," *IEEE Signal Process. Mag.*, vol. 28, no. 1, pp. 33–42, Jan. 2011.

[3] T. Hermann, A. Hunt, and J.G. Neuhoff, *The Sonification Handbook*, Logos Publishing House, 2011.

[4] J. Blauert, *Spatial Hearing - The Psychophysics of Human Sound Localization*, MIT Press, Cambridge, rev. edition, 1997.

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

[6] M. Geier, J. Ahrens, and S. Spors, "The soundscape renderer: A unified spatial audio reproduction framework for arbitrary rendering methods," in *124th AES Conv.*, 2008.

[7] M. Vorländer, *Auralization (RWTHedition)*, Springer Berlin Heidelberg, 2007.

[8] S. Malik, J. Fligge, and G. Enzner, "Continuous HRTF acquisition vs. HRTF interpolation for binaural rendering of dynamical auditory virtual environments," *ITG-Fachbericht-Sprachkommunikation*, 2010.

[9] K. Sunder, J. He, E. L. Tan, and W. S. Gan, "Natural sound rendering for headphones: Integration of signal processing techniques," *IEEE Signal Process. Mag.*, vol. 32, no. 2, pp. 100–113, Mar. 2015.

[10] F. Rumsey, *Spatial Audio*, Music Technology Series. Focal Press, 2001.

[11] S. Carlile, *Virtual Auditory Space: Generation and Applications*, Neuroscience Intelligence Unit. Landes Bioscience, 1996.

[12] D. S. Brungart, B. D. Simpson, and R. L. Mckinley, "The interaction between head-tracker latency, source duration, and response time in the localization of virtual sound sources," in *Proc. of the Int. Conf. on Auditory Display (ICAD)*, 2004.

[13] A. Lindau, "The perception of system latency in dynamic binaural synthesis," *Proc. of 35th DAGA*, pp. 1063–1066, 2009.

[14] D. S. Brungart, B. D. Simpson, and A. J. Kordik, "The detectability of headtracker latency in virtual audio displays," in *Proc. of the 11th Int. Conf. on Auditory Display (ICAD)*, 2005.

[15] A. Andreopoulou, D. R. Begault, and B. F. G. Katz, "Interlaboratory round robin HRTF measurement comparison," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 5, pp. 895–906, Aug. 2015.

[16] V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano, "The CIPIC HRTF database," in *Proceedings of the 2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics*, 2001.

[17] W. G. Gardner and K. D. Martin, "HRTF measurements of a KEMAR," *Acoustical Society of America Journal*, vol. 97, pp. 3907–3908, 1995.

[18] J. Blauert, M. Brueggen, K. Hartung, A. W. Bronkhorst, R. Drullmann, G. Reynaud, L. Pellieux, W. Krebber, and R. Sottek, "The AUDIS catalog of human HRTFs," in *16th Int. Congress of Acoustics*, 1988.

[19] N. Gupta, A. Barreto, M. Joshi, and J. C. Agudelo, "HRTF database at FIU DSP Lab," in *2010 IEEE Int. Conf. on Acoustics, Speech and Signal Proc.*, Mar. 2010, pp. 169–172.

[20] G. Enzner, "Analysis and optimal control of lms-type adaptive filtering for continuous-azimuth acquisition of head related impulse responses," in *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, Mar. 2008, pp. 393–396.

[21] G. Enzner, "3D-continuous-azimuth acquisition of head-related impulse responses using multi-channel adaptive filtering," in *2009 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, Oct. 2009.

[22] J. Blauert, *The Technology of Binaural Listening*, Modern Acoustics and Signal Processing. Springer Berlin Heidelberg, 2013.

[23] P. Faundez Hoffmann and H. Møller, "Audibility of spectral switching in head-related transfer functions," in *Audio Engineering Society Conv. 119*, Oct. 2005.

[24] M. Otani and T. Hirahara, "A dynamic virtual auditory display: Its design, performance, and problems in HRTF switching," in *The Japan-China Joint Conf. of Acoustics*, 2007.

[25] A. Kudo, H. Hokari, and S. Shimada, "A study on switching of the transfer functions focusing on sound quality," *Acoustical Science and Technology*, vol. 26, no. 3, pp. 267–278, 2005.

[26] V. Valimaki, A. Franck, J. Ramo, H. Gamper, and L. Savioja, "Assisted listening using a headset: Enhancing audio perception in real, augmented, and virtual environments," *IEEE Signal Process. Mag.*, vol. 32, no. 2, pp. 92–99, March 2015.

[27] J. W. Scarpaci, H. S. Solburn, and J. A. White, "A system for real-time virtual auditory space," in *11th Meeting of the Int. Conf. on Auditory Display*, July 2005.

[28] B. Xie, *Head-Related Transfer Function and Virtual Auditory Display*, A Title in J. Ross Publishing's Acoustics: Information and Communication Series. J. Ross Publishing, 2013.

[29] H. Kuttruff, *Akustik: Eine Einführung*, Hirzel, 2004.

[30] L.E. Kinsler, *Fundamentals of Acoustics*, Wiley, 2000.

[31] E. König, "Pitch discrimination and age," *Acta Oto-Laryngologica*, vol. 48, no. 5-6, pp. 475–489, 1957.

[32] O. H. Meurmann, "The difference limen of frequency in tests of auditory function," *Acta Oto-Laryngologica*, vol. 43, no. sup118, pp. 144–155, 1954.

[33] E. Jan, P. Svaizer, and J. L. Flanagan, "Matched-filter processing of microphone array for spatial volume selectivity," in *Circuits and Systems, 1995. ISCAS '95., 1995 IEEE International Symposium on*, Apr. 1995, vol. 2, pp. 1460–1463 vol.2.

[34] J. H. DiBiase, H. F. Silverman, and M. S. Brandstein, "Robust localization in reverberant rooms," in *Microphone Arrays*, pp. 157–180. Springer, 2001.

[35] M. Zohourian, G. Enzner, and R. Martin, "Binaural speaker localization integrated into an adaptive beamformer for hearing aids," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 26, no. 3, pp. 515–528, March 2018.