ADVANCED SYSTEM OPTIONS FOR BINAURAL RENDERING OF AMBISONIC FORMAT

Gerald Enzner, Michael Weinert

Institute of Communication Acoustics
Ruhr-Universität Bochum
D-44780 Bochum, Germany
gerald.enzner@rub.de

Stefan Abeling, Jan-Mark Batke, Peter Jax

Technicolor Research & Innovation
Karl-Wiechert-Allee 74
D-30625 Hannover, Germany
peter.jax@technicolor.com

ABSTRACT

Ambisonics uses a well-respected soundfield representation and thus may become a standard in soundfield transmission, storage, and reproduction. Then we will need signal processing options in order to decode and apply the Ambisonic format universally in various applications from loudspeaker to headphone reproduction. More specifically, the use of Ambisonics is envisioned also in mobile applications where headphone reproduction is clearly preferred. In this context, we present two different system options for the advanced implementation of binaural soundfield rendering. Both options utilize the concept of virtual loudspeakers and subsequent binaural rendering via HRTF. In order to achieve spatial realism, one option relies on adaptive soundfield rotation in the Ambisonic domain to mimic a corresponding head rotation of the listener in the given soundfield. The other system option relies on a new continuous-azimuth HRTF format for direct representation of head rotations within the virtual loudspeaker setup. For both system options, we will discuss the pros and cons regarding the implementation. We conclude with a comparison of the resulting localization accuracies in listening tests.

Index Terms— Spatial audio, binaural technology, perception

1. INTRODUCTION AND RELATION TO PRIOR WORK

The Ambisonic encoding of soundfields [1] via the truncated spherical harmonics expansion [2] provides flexible means for storage, transmission, and reproduction of the soundfields in a remote environment [3]. By its universal mathematical formulation, cf. Sec. 2.1, the Ambisonic representation in particular achieves decoupling between recording and reproduction setups. Therefore, the awareness for the specific loudspeaker playback configuration can be limited to the decoder side. Moreover, the Ambisonic representation allows a variety of efficient soundfield modifications, such as rotation, holophonic translation, and scalability of the spatial resolution.

Due to its flexibility, the Ambisonic soundfield format may turn into a standard for 3D audio equipment in applications such as surround sound for professional and home entertainment, computer games, or teleconferencing with spatial separation of multiple participants. In all cases, it will be required to achieve Ambisonic decoding not only for particular loudspeaker arrangements, but also for mobile terminals with personalized headphone reproduction. A good example in that respect is the previous MPEG Surround specification and its large set of binaural rendering extensions [4]. In order to achieve a stable sound image in mobile-headphone applications, it will be further necessary to utilize head-tracking information to unlock the reproduced soundfield from head movements [5].

An intuitive approach to Ambisonic headphone-decoding relies on the two-stage concept of 1) decoding to a set of virtual loudspeakers and 2) filtering each loudspeaker signal with the respective pair of head-related transfer functions (HRTFs) [3, 6], cf. Sec. 2.2. At first sight, this introduces more degrees of freedom than for real-loudspeaker decoding, e.g., regarding the number of virtual loudspeakers and their unrestricted placement in the virtual environment. However, considering the typical limitation to finite order N in the context of Ambisonic encoding, e.g., when using spherical microphone arrays [7, 8], the related number of $(N+1)^2$ Ambisonic coefficients then naturally invokes a configuration with just $(N+1)^2$ real or virtual loudspeakers (unless we aim at interpolated reproduction with more loudspeakers). Another limitation is given by the conventional spatial sampling of (individualized or non-individualized) HRTF tables in the typical order of 5 to 10 deg azimuth- and 10 to 20 deg elevation-spacing on the auditory sphere [9]. In conjunction with natural head-movements, head-tracking, and HRTF adjustment w.r.t. the virtual loudspeakers, a mismatch of loudspeaker placement and spatially discrete HRTF support will occur. This would invoke sophisticated considerations regarding HRTF interpolation or cause a degradation of spatial sound fidelity otherwise.

The literature has been presenting essentially two options to decode Ambisonic format for binaural playback controlled by head-tracking. Regarding the first "soundfield-oriented" option, early papers highlight the equivalence between a rotation of the listener's head or the acoustic environment, i.e., the soundfield, and on this basis apply "trivial" rotation matrices to B-format (first-order Ambisonics) soundfield coefficients [10, 11]. A similar system using "simple" matrices was reported a decade later for higher-order Ambisonics (HOA) decoding and binaural playback [12], but the mathematical operations for soundfield rotation were not included. In Sec. 3.1 of our paper, we thus revisit the "soundfield-oriented" system option on explicit mathematical grounds.

The second and more "listener-bound" option performs plane-wave decomposition [13] of an incident soundfield [14], spherical-harmonic-based beamforming [15], or mode-matching [16, 3, 17] to decode an acoustic recording into components arriving from various directions. Those components are then weighted with plane-wave HRTFs and summed up to yield the reproduction signal for each ear. Based on head-tracking data, the HRTFs are retrieved dynamically via nearest-neighbor search from a database or generated on the fly via interpolation for the respective virtual loudspeaker directions. In Sec. 3.2 of our paper, we advance this "listener-bound" system option by applying individualized and quasi-continuous HRTF data [18] to avoid any HRTF related errors in binaural decoding.

Sec. 4 finally presents an assessment of the advanced system options by means of listening tests. Relationship of actual and perceived sound source directions is depicted. For reference, we include ideal (i.e., spatially-continuous) HRTF-based binaural rendering.

2. SIGNALS AND SYSTEMS OF 3D SOUND

Here, we briefly define our HOA-based soundfield and HRTF-based listener representations. Afterwards in Secs. 3.1 and 3.2, the goal will be to connect the incident soundfield and the binaural receiver via head-tracking to achieve immersive binaural rendering.

2.1. Ambisonic Representation of the Soundfield

Let us consider the acoustic configuration in spherical (r, ϑ, φ)-coordinates in Fig. 1, where elevation ϑ is not depicted explicitly. According to the acoustic wave-equation, we represent a narrow-band incident sound-pressure field $p(r, \vartheta, \varphi)$ at the position of the listener in terms of the Fourier-Bessel series [2, 16]

$$p(r, \vartheta, \varphi, \omega) = \sum_{n=0}^{N} \sum_{m=-n}^{n} A_{mn}(\omega) j_n(kr) Y_n^m(\vartheta, \varphi) \ , \quad (1)$$

where $j_n(kr)$ denotes the spherical Bessel functions, $Y_n^m(\vartheta, \varphi) = f(n, m) L_n^m(\cos\vartheta) \mathrm{e}^{im\varphi}$ the spherical harmonics based on the associated Legendre polynomials $L_n^m(\cos\vartheta)$ and a function $f(n, m)$, $\omega = 2\pi f$ the acoustic frequency, $k = \omega/c$ the wavenumber, c the speed of sound, N the expansion order, and A_{mn} the independent coefficients to represent the actual soundfield.

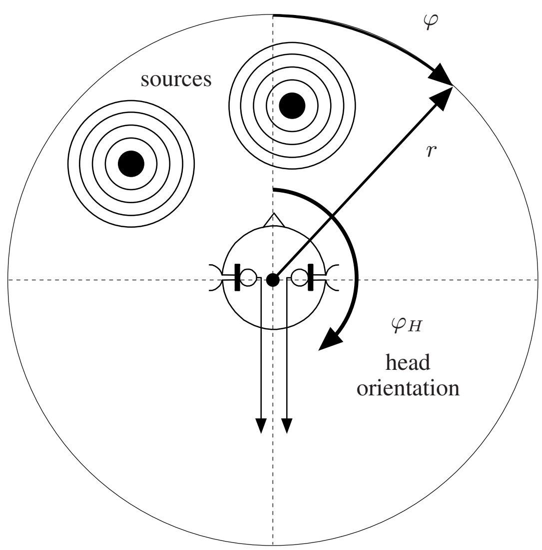


Fig. 1. Natural sound sources in a user-centered coordinate system.

The concept of the "incident" or "incoming" or "through-going" soundfield means that the effect of scattering off the listener or off another soundfield measurement device is not represented by this expansion type. The effect of scattering would be in fact undesirable in this soundfield representation, since later we will take the scattering off the individual into account via free-field HRTF processing of the soundfield. A scatter-free soundfield representation is analytically known for incident plane waves or spherical waves of any order [2], while sophisticated microphone arrays are used to obtain a scatter-free representation of natural soundfields [7, 8, 14]. For the latter, however, the order is typically limited to small numbers.

A further parameter defined by Fig. 1 is the head-orientation φ_H, which will be considered more specifically in Secs. 3.1 and 3.2.

2.2. HRTF Representation of the Listener

Fig. 2 depicts an acoustic listening environment, real or virtual, with discrete sound sources in the form of loudspeakers at exemplary locations $\Omega_{LS,i} = (\vartheta_{LS,i}, \varphi_{LS,i})$ and distances $r \to \infty$, i.e., in the far-field of the listener. Head-related impulse responses (HRIRs) $h_\kappa^{l/r}(\Omega)$ are then defined by the convolutive relationship

$$y_i^{l/r}(k) \mid_{\varphi_H=0} = \sum_{\kappa=0}^{K} x_{\Omega_{LS,i}}(k-\kappa) h_\kappa^{l/r}(\Omega_{LS,i}) \quad (2)$$

between a single source signal $x_{\Omega_{LS,i}}(k)$ and the respective responses $y_i^{l/r}(k) \mid_{\varphi_H=0}$ at the left and right ears, where default head-rotation $\varphi_H = 0$ was assumed.

Those far-field HRIRs $h_\kappa^{l/r}(\Omega)$ are plenacoustic functions, i.e., they naturally extend over all possible solid angles $\Omega = (\vartheta, \varphi)$ in the auditory sphere. For a given loudspeaker direction $\Omega_{LS,i}$ and arbitrary head-orientation $\Omega_H = (\vartheta_H, \varphi_H)$, the effective HRIR direction to this loudspeaker then obviously amounts to $\Omega_{\mathrm{eff},i} = \Omega_{LS,i} - \Omega_H$. In order to apply HRIR/HRTF processing in a virtual environment, including the desired Ambisonic headphone-decoding, the HRIRs have to be determined in an anechoic acoustic lab environment beforehand and stored in a suitable database format, e.g., [19, 6].

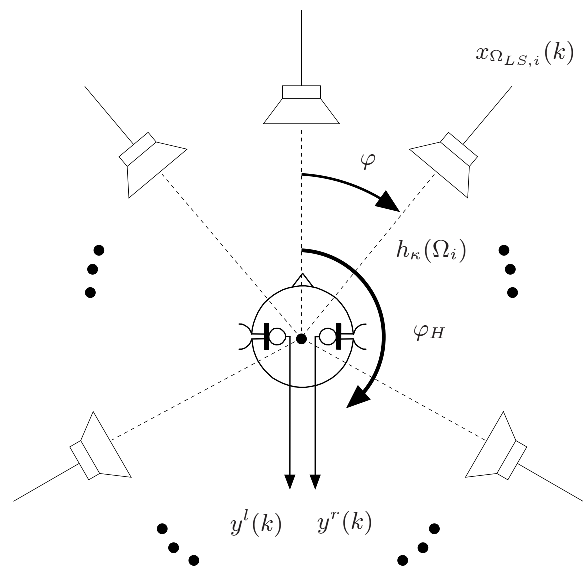


Fig. 2. Listener in a real or virtual multichannel loudspeaker system.

3. SYSTEMS FOR HEAD-TRACKED BINAURALIZATION

Our two system options for binaural rendering of Ambisonic format will share several state-of-the-art components:

- Advanced headphone-mounted head-tracking to deliver φ_H.
- Ambisonic decoding of $(N{+}1)^2$ original or manipulated HOA coefficients $A_{mn}(\omega)$ into $(N{+}1)^2$ far-field loudspeaker input signals $x_{\Omega_{LS,i}}(k)$. This operation can be performed in various ways, e.g., via far-field mode-matching [3, 17] or explicit plane-wave decomposition [14].
- Individual HRIR convolution $(*)$ for left and right ear. Summation (Σ) over all virtual loudspeakers.

More precise specification of the components will be reported in Sec. 4. Right here, we describe two variants of principle usage.

3.1. "Soundfield-Oriented" System Option

Based on the equivalence of head-rotation and opposite soundfield-rotation [11], our "soundfield-oriented" system option in Fig. 3 uses the head-rotation φ_H as an input to a soundfield rotation module in the Ambisonic coefficients domain. This key element of the system essentially exploits the spatial interpolation property related to the spatially-continuous Ambisonic format. The rotated soundfield is represented by the Ambisonic coefficients B_{mn} which are then decoded to virtual loudspeaker positions $\Omega_{LS,i}$. The resulting loudspeaker signals $x_{\Omega_{LS,i}}(k)$ are binaurally rendered for headphone reproduction using a small, fixed, and spatially discrete HRIR set corresponding to $\Omega_{LS,i}$, as if no head-rotation has taken place, i.e.,

$$y^{l/r}(k) = \sum_{i}^{(N+1)^2} y_i^{l/r}(k)\,|_{\varphi_H=0} \quad . \tag{3}$$

At this point, we shall not leave the concept of soundfield manipulation in the Ambisonic domain to words alone. We therefore briefly include mathematical representation for our example of soundfield rotation by angle φ_H along the azimuth coordinate. To this end, we rely on the Fourier-Bessel series in (1) to express the soundfield coefficients A_{mn} as a spatial Fourier analysis of the given soundfield $p(r,\vartheta,\varphi,\omega)$, e.g. [2],

$$A_{mn}(\omega) = j_n^{-1}(kr)\int_{\Omega} p(r,\vartheta,\varphi,\omega)Y_n^m(\vartheta,\varphi)^*\mathrm{d}\Omega\;, \tag{4}$$

where $\mathrm{d}\Omega = \sin\vartheta\mathrm{d}\vartheta\mathrm{d}\varphi$. The rotated soundfield $p(r,\vartheta,\varphi+\varphi_H,\omega)$ is thus represented, via substitution of Ω, by soundfield coefficients

$$\begin{aligned}
B_{mn}(\omega) &= j_n^{-1}(kr)\int_{\Omega} p(r,\vartheta,\varphi+\varphi_H,\omega)Y_n^m(\vartheta,\varphi)^*\mathrm{d}\Omega \\
&= j_n^{-1}(kr)\int_{\Omega'} p(r,\vartheta,\varphi',\omega)Y_n^m(\vartheta,\varphi'-\varphi_H)^*\mathrm{d}\Omega' \\
&= j_n^{-1}(kr)\int_{\Omega'} p(r,\vartheta,\varphi',\omega)Y_n^m(\vartheta,\varphi')^*\mathrm{e}^{im\varphi_H}\mathrm{d}\Omega' \\
&= \mathrm{e}^{im\varphi_H}A_{mn}(\omega)\;,
\end{aligned} \tag{5}$$

where the auxiliary operator $\mathrm{d}\Omega' = \sin\vartheta\mathrm{d}\vartheta\mathrm{d}\varphi'$ has been applied.

The required manipulation to achieve the desired azimuth-rotation is obviously straightforward in the Ambisonic domain. The above derivation proved a simple law-of-modulation which is in line with the theory of 1-dimensional linear systems, e.g., [20]. The factor $\mathrm{e}^{im\varphi_H}$ is even independent of the elevation modes n. It should be noted, however, that a 2-dimensional Ω_H rotation creates far more complicated expressions with highly nonlinear interdependencies of n and m. Here we just refer to a recent presentation [21] which in turn refers to seminal work by Wigner [22]. Its complexity does not seem to be attractive from the engineering perspective. Numerical solutions for this case are, however, beyond the scope of our paper.

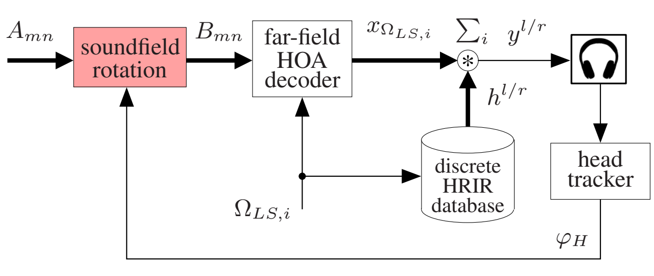


Fig. 3. System based on Ambisonic-format soundfield manipulation.

3.2. "Listener-Bound" or "Continuous-HRIR" System Option

The newly proposed implementation in Fig. 4 directly utilizes the effective HRIRs $\Omega_{\mathrm{eff},i} = \Omega_{LS,i} - \Omega_H$ for given loudspeaker positions to accommodate for any head-orientation $\Omega_H = (0,\varphi_H)$ by the listener. After conventional Ambisonic decoding to virtual loudspeakers, the actual binaural rendering is performed efficiently via HRIR convolution in time-domain and summation over all loudspeakers:

$$y^{l/r}(k) = \sum_{i}^{(N+1)^2}\sum_{\kappa=0}^{K} x_{\Omega_{LS,i}}(k-\kappa)h_\kappa^{l/r}(\Omega_{LS,i}-\Omega_H)\;. \tag{6}$$

Considering this result as a function of the head-orientation Ω_H, this formula actually depicts similarity with discrete spatial convolution on the far-field loudspeaker sphere. This in turn indicates a chance of utilizing the spherical convolution theorem [23] to achieve unification or at least harmonization with the "soundfield-oriented" system in the Ambisonic domain. This, too, is beyond our scope here.

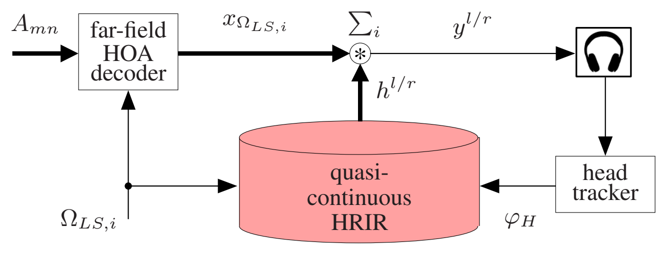


Fig. 4. New "direct" system based on sophisticated HRIR database.

While the direct "listener-bound" implementation in Fig. 4 usually requires interpolation of HRIR tables, we rely on a new and quasi-continuous HRIR format. Using [18], very-high-resolution HRIRs are measured conveniently within 1-2 minutes for each individual. This HRIR measurement applies a slow but continuous rotation of the subject of interest during acquisition. Adaptive filters then extract the rotating HRIR at any azimuth from the in-ear recordings. For backwards compatibility with conventional discrete HRIR databases, our sampling of the spatially-continuous HRIR, however, requires huge memory depending on the desired resolution [24].

4. IMPLEMENTATION AND EVALUATION

The primary interest of our paper is to compare the two system options, i.e., the "soundfield rotation" and "continuous HRIR" options from Secs. 3.1 and 3.2, respectively. A "reference" system is further included which performs direct HRIR-rendering of the respective sound stimulus, i.e., no Ambisonic format is involved in this idealized reference configuration. In the following, we briefly mention various implementation parameters that remain fixed in our experiments. Afterward we report and interpret the localization results obtained from our informal listening tests.

4.1. System Implementation and Configuration

Our system input are plane waves arriving from $\Omega_l = (\vartheta_l,\varphi_l)$. It is thus represented by $A_{mn} = 4\pi i^n Y_n^m(\vartheta_l,\varphi_l)^*$, e.g. [2, Eq. 6.175], using a realistic and yet ambitious order $N = 6$. The actual stimuli to be conveyed by the plane waves are reported below. HOA decoding is implemented via classical mode-matching, e.g. [17], and we decode to $(N+1)^2 = 49$ almost uniformly distributed virtual loudspeaker positions according to Fliege grids [25]. This amounts to roughly 30 deg angular spacing between any two loudspeakers.

In all experiments, we make use of individualized HRIR data according to [18]. For the "soundfield rotation" system, this quasi-continuous HRIR data (here: 0.1 deg azimuth spacing) is decimated to a much smaller database that matches the 49 discrete loudspeaker positions. In all cases, the HRIR databases are represented in the *OpenDAFF* file format [26, 24], which was found to be very effective for realtime supply of individualized and changing HRIRs.

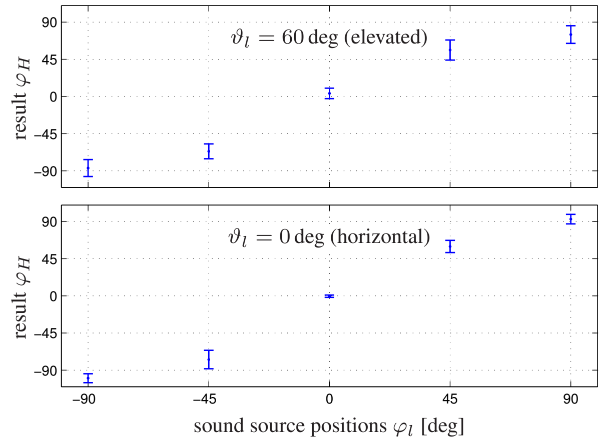
All systems, including the "reference", use the *Polhemus Fastrak* to supply the respective system with the actual head-orientation. The resulting binaural signal is finally presented to 6 untrained normal-hearing listeners via *Beyerdynamic DT 770 M* headphones. No further calibration or headphone equalization is applied.

4.2. Source Localization (Source Positioning) Test

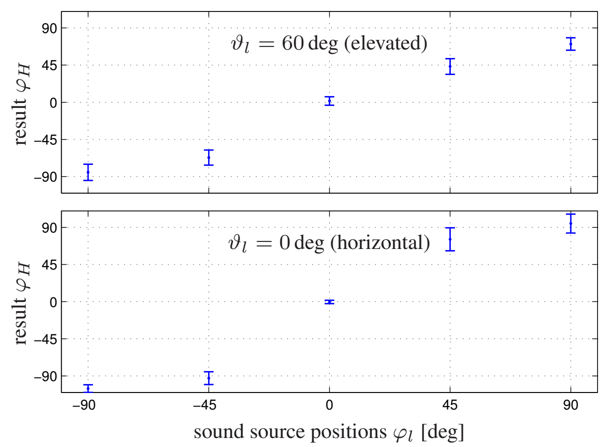
Source localization tests require sophisticated pointing methods for the listener to indicate the perceived sound source direction, e.g., [27, 28] and numerous references therein. Regarding the established head-pointing methods, i.e., looking into the source direction and then exploiting the head-tracking result φ_H, we object the effect of eventually listening from the front direction in all cases. We therefore suggest a somewhat unusual head-pointing method:

A first sound source (here: *EBU SQAM #53 speech*, 44.1 kHz) is placed into a fixed reference position in the virtual auditory space (here: $\varphi = 0$ and different elevations $\vartheta = \vartheta_l$). A second sound source (here: *EBU SQAM #02 pink noise*, 44.1 kHz) is placed into various directions of interest (here: $\varphi_l \in \{-90, -45, 0, 45, 90\,\mathrm{deg}\}$ and $\vartheta_l \in \{0, 60\,\mathrm{deg}\}$). It is noteworthy in our context that only a subset of these directions can coincide with our virtual loudspeaker positions. The second source is locked with the head and, starting from $\varphi_H = 0$, the task of the listener is to match the two sources via his/her head rotation. In this way, the direction of interest is preserved during listening and, upon decision of the listener, the head-tracker output φ_H is sampled to indicate this direction. Three shuffled repetitions are considered for each direction of interest.

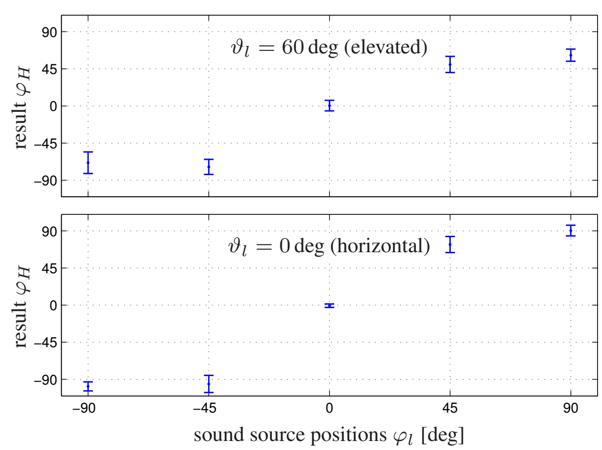
Fig. 5 depicts the mean localization results φ_H over all subjects and the corresponding standard deviations for each direction Ω_l under test. Let us first consider the horizontal plane. Already for the "reference" system, we notice a characteristic deviation between sound source localization and true sound source position, i.e., the chosen HRIR direction. Especially for $|\varphi_l| = 45$ deg, the perception is obviously "pulled" a bit towards $|\varphi_H| = 90$ deg. This deviation between original and perceived sound source location confirms early studies of free-field human sound localization [29], headphone simulation of free-field listening [30], and in particular previous work using speech stimuli [31, 32]. The observed deviation is rooted in two unfortunate phenomena, i.e., 1) the cones-of-confusion known from spatial hearing [19] and 2) the inside-head localization with headphones, which effectively turns into lateralization [33, 32]. For both HOA systems of finite order $N = 6$, i.e., "soundfield rotation" and "continuous HRIR", the observed characteristic is preserved up to only a bit more pronunciation. This result proves the suitability of Ambisonic rendering by virtual loudspeakers and HRIRs. At higher elevation, interestingly, the characteristic is different in that the source at $|\varphi_l| = 90$ deg is consistently pulled a bit towards the front of the listener for all systems, including the reference.



(a) Ideal HRTF-based binaural rendering ("the reference").



(b) The "soundfield-oriented" or "soundfield-rotation" system.



(c) The "listener-bound" or "continuous-HRIR" system.

Fig. 5. Mean sound localization results and standard deviations [34].

5. CONCLUSIONS

Advances in HRTF-technology have enabled a new degree of flexibility in the implementation of virtual loudspeaker concepts for binaural rendering, especially, the seamless representation of head-movements via continuous HRTF. In our study, the previously suggested "soundfield rotation" and the newly proposed "continuous HRTF" system were compared and turned out to be similar in terms of localization accuracy. The actual choice then depends on other criteria such as the specific hardware/software architecture, i.e., specifically the resources on an application specific platform. Low memory requirements, e.g., would still invoke "soundfield rotation".

6. REFERENCES

[1] M. Gerzon, "Ambisonics in multichannel broadcasting and video," *J. Audio Enggr. Soc.*, vol. 33, no. 11, pp. 859–871, Nov. 1985.

[2] E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography*, Academic Press, San Diego, London, 1999.

[3] J. Daniel, "Spatial sound encoding including near field effect: Introducing distance coding filters and a viable, new Ambisonic format," in *AES 23rd Intl. Conf.*, Copenhagen, May 2003.

[4] J. Breebart, J. Herre, L. Villemoes, C. Jin, K. Kjörling, J. Plogsties, and J. Koppens, "Multichannel goes mobile: MPEG Surround binaural rendering," in *AES 29th Intl. Conf.*, Seoul, Sep. 2006.

[5] C. Pörschmann, "3D audio in mobile communication devices: Methods for mobile head-tracking," *Journal of Virtual Reality and Broadcasting*, vol. 4, no. 13, 2007.

[6] R. Nicol, *Binaural Technology*, AES Monograph, New York, 2010.

[7] J. Meyer and G. Elko, "A spherical microphone array for spatial sound recordings," *J. Acoust. Soc. Am.*, vol. 111, pp. 2346–2356, 2002.

[8] T. Abhayapala and D. Ward, "Theory and design of high order sound field microphones using spherical microphone arrays," in *Proc. IEEE Intl. Conf. Acoust., Speech, Signal Proc. (ICASSP)*, Orlando, May 2002.

[9] F. Wightman and D. Kistler, "Measurement and validation of human HRTFs for use in research," *Acta Acustica united with Acustica*, vol. 91, pp. 429–439, Oct. 2005.

[10] D. Malham, "3-D sound for virtual reality systems using Ambisonic techniques," in *3rd Annual Conf. Virtual Reality*, London, UK, April 1993.

[11] A. McKeag and D. McGrath, "Sound field format to binaural decoder with head-tracking," in *AES Conv. 6r*, Sidney, Aug. 1996, p. Paper No. 4302.

[12] M. Noisternig, A. Sontacchi, T. Musil, and R. Höldrich, "A 3D Ambisonic based binaural sound reproduction system," in *AES 24th Intl. Conf. on Multichannel Audio*, Banff, Canada, June 2003.

[13] D. Zotkin, R. Duraiswami, and N. Gumerov, "Plane-wave decomposition of acoustical scenes via spherical and cylindrical microphone arrays," *IEEE Trans. Audio, Speech and Language Process.*, vol. 18, no. 1, pp. 2–16, Jan. 2010.

[14] R. Duraiswami, D. Zotkin, Z. Li, E. Grassi, N. Gumerov, and L. Davis, "High order spatial audio capture and its binaural head-tracked playback over headphones with HRTF cues," in *AES 119th Conv.*, New York, Oct. 2005.

[15] A. O'Donovan, D. Zotkin, and R. Duraiswami, "Spherical microphone array based immersive audio scene rendering," in *Proc. of the 14th Intl. Conf. on Auditory Displays*, Paris, June 2008.

[16] J. Daniel, R. Nicol, and S. Moreau, "Further Investigations of High Order Ambisonics and Wavefield Synthesis for Holophonic Sound Imaging," in *114th AES Conv.*, Amsterdam, March 2003.

[17] M. Poletti, "Three-dimensional surround sound systems based on spherical harmonics," *J. Audio Enggr. Soc.*, vol. 53, no. 11, pp. 1004–1025, Nov. 2005.

[18] G. Enzner, "3D-continuous-azimuth acquisition of head-related impulse responses using multi-channel adaptive filtering," in *Proc. IEEE Workshop on Applications of Signal Process. to Audio and Acoust. (WASPAA)*, New Paltz, NY, Oct. 2009.

[19] J. Blauert, *Spatial Hearing*, The MIT Press, Cambridge, MA, 1997.

[20] T. Kailath, *Linear Systems*, Prentice Hall, Inc., 1980.

[21] Z. Gimbutas and L. Greengard, "A fast and stable method for rotating spherical harmonic expansions," *J. Comp. Physics*, vol. 228, no. 16, pp. 5621–5627, Sep. 2009.

[22] E. Wigner, *Gruppentheorie und ihre Anwendungen auf die Quantenmechanik der Atomspektren*, Vieweg Verlag, Braunschweig, 1931.

[23] J. Driscoll and D. Healy, "Computing Fourier Transforms and Convolutions on the 2-Sphere," *Advances in Applied Mathematics*, vol. 15, no. 2, pp. 202–250, June 1994.

[24] M. Weinert, G. Enzner, J.-M. Batke, P. Jax, and C. Antweiler, "Komfortable Messung und Bereitstellung individueller kopfbezogener Impulsantworten als OpenDAFF," in *Proc. of Deutsche Jahrestagung für Akustik (DAGA)*, Darmstadt, Germany, March 2012.

[25] J. Fliege and U. Maier, "The distribution of points on the sphere and corresponding cubature formulae," *IMA Journal of Numerical Analysis*, vol. 19, no. 2, pp. 317–334, 1999.

[26] F. Wefers, "OpenDAFF – A Free, Open-Source Software Package for Directional Audio Data," in *Proc. of Deutsche Jahrestagung für Akustik (DAGA)*, Berlin, Germany, March 2010.

[27] T. Djelani, C. Pörschmann, J. Sarhage, and J. Blauert, "An interactive virtual-environment generator for psychoacoustic research II: Collection of head-related impulse responses and evaluation of auditory localization," *Acta Acustica united with Acustica*, vol. 86, no. 6, pp. 1046–1053, Nov./Dec. 2000.

[28] P. Majdak, B. Laback, M. Goupell, and M. Mihocic, "The accuracy of localizing virtual sound sources: Effects of pointing method and visual environment," in *124th AES Conv.*, Amsterdam, May 2008.

[29] S. Oldfield and S. Parker, "Acuity of sound localisation: a topography of auditory space. I. Normal hearing conditions," *Perception*, vol. 13, no. 5, pp. 581–600, 1984.

[30] F. Wightman and D. Kistler, "Headphone simulation of free-field listening II: Psychophysical validation," *J. Acoust. Soc. Am.*, vol. 85, pp. 868–878, 1989.

[31] D. Begault and E. Wenzel, "Headphone localization of speech," *Human Factors*, vol. 35, no. 2, pp. 361–376, 1993.

[32] D. Begault, *3-D Sound for Virtual Reality and Multimedia*, National Aeronautics and Space Administration, 2000.

[33] G. Plenge, "On the differences between localization and lateralization," *J. Acoust. Soc. Am.*, vol. 56, no. 3, pp. 944–951, Sep. 1974.

[34] J. Spille, "Participation and comprehensive feedback during development of the listening test," Personal Communication, Aug./Sep./Oct. 2012.