Nearfield binaural synthesis and ambisonics

Dylan Menzies[a] and Marwan Al-Akaidi
Department of Computer Sciences and Engineering, De Montfort University, Leicester, Leicestershire LE1 6RS United Kingdom

(Received 20 June 2006; revised 22 December 2006; accepted 23 December 2006)

Ambisonic encodings can be rendered binaurally, as well as for speaker arrays. This process is developed for general high-order Ambisonic encodings of soundfields containing near as well as far sources. For sufficently near sources an error is identified, resulting from the limited field of validity of the freefield harmonic expansion. A modified expansion is derived that can render such sources correctly. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2434761]

PACS number(s): 43.60.Sx, 43.66.Pn, 43.66.Qp [EJS] Pages: 1559–1563

I. INTRODUCTION

A. Ambisonic encoding

Ambisonics is a methodology developed for encoding soundfields, and decoding them onto speaker arrays, (Gerzon, 1985, 1992). Initially it was used only to first order, with four signals that encode a full sphere of sound around a central listener. More recently Ambisonics has been employed at higher orders, whereby it is possible to not only increase the angular resolution of distant sources, but also extend the listening region and recreate accurately the soundfield from nearfield sources, (Daniel, 2003). We shall refer informally to an encoding of any order as *B format*, borrowing the original terminology for first order. Using high-order encodings, the listener receives distance cues about near sources exactly as they would for the real soundfield, because the soundfield around the listener can be reconstructed arbitrarily well.

The ambisonic encoded signals are defined by a spherical harmonic expansion of the freefield. Although our discussion does not depend on a particular representation, for definiteness we use signals, $B^{\sigma}_{mn}(k)$, defined with the real-valued *N3D* spherical harmonics, (Daniel, 2000),

$$\Psi(\boldsymbol{r},k)=\sum_{m} i^{m} j_{m}(kr)\sum_{n,\sigma} Y^{\sigma}_{mn}(\theta,\delta)B^{\sigma}_{mn}(k), \tag{1}$$

where

$$Y^{\sigma}_{mn}(\theta,\delta)=\sqrt{2m+1}\,\widetilde{P}_{mn}(\sin\delta)\times\begin{cases}\cos n\theta & \text{if } \sigma=+1\\ \sin n\theta & \text{if } \sigma=-1\end{cases} \tag{2}$$

$$\widetilde{P}_{mn}(\sin\delta)=\sqrt{(2-\delta_{0,n})\frac{(m-n)!}{(m+n)!}}P_{mn}(\sin\delta). \tag{3}$$

For $n=0$, σ only takes the value $+1$. θ here measures the angle around the coordinate symmetry axis, and $\pi/2-\delta$ is the angle between the axis and the coordinate direction, so that δ would normally be called the elevation.

From here on we shall use a slightly simplified notation that removes the need for σ by extending n to negative values as used in the standard complex set

$$Y_{mn}=\begin{cases}Y^{+1}_{mn} & \text{if } n\geq 0\\ Y^{-1}_{m|n|} & \text{if } n<0\end{cases}. \tag{4}$$

Similarly the encoded signals become $B_{mn}(k)$.

B. Conversion to binaural

In a binaural rendering system the listener is presented with one signal to each ear canal direct. Binaural signals can be derived from an ambisonic encoding, using head related transfer functions (HRTFs), as described below. The ambisonic encoding is easily rotated, which facilitates compensation for head movement.

Conversion to binaural can be achieved approximately by summing speaker array feeds that are each filtered by a HRTF matching the speaker position, (Jot and Wardle, 1998; McKeag and McGrath, 1996). Figure 1 illustrates the signal flow in this process.

A natural extension of this idea to an exact method for binaural signals is to transform the encoded soundfield into a plane wave expansion, and weight each component plane wave by the plane wave HRTF matching its direction and frequency, (Duraiswami *et al.*, 2005; Menzies, 2002). The process can be applied to high-order ambisonic encodings containing sources at various distances. A straightforward binaural approach would require HRTF sets for each source distance, however decoding the high-order signal requires only the plane wave HRTF set. This is not too surprising, as the HRTF sets are defined within the contraints of the wave equation, and so are all related. There is another less obvious advantage, which is that complex sources can be conveniently converted to binaural via high-order B format, as will be demonstrated in a future article. A single nearfield HRTF set cannot be applied in a simple way to a complex source description to yield the required binaural signals. Figure 2 depicts an overview of the encoding process from soundfield to binaural using a plane wave expansion.

The process is exact in the farfield, but as explained later, there is a subtle source of error which can affect near

[a]Author to whom correspondence should be addressed. Electronic mail: dylan@dmu.ac.uk

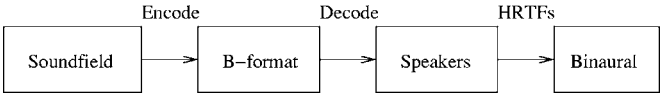

FIG. 1. Encoding a soundfield to binaural via virtual speakers.

sources. First we detail the steps to generate binaural signals from high-order B format using a plane wave expansion.

We aim to represent a source-free region by an expansion in plane waves, known as a Herglotz expansion, with coefficients $\mu(\boldsymbol{s},k)$ defined over unit vectors s, so that

$$\Psi(\boldsymbol{r},k)=\frac{1}{4\pi}\int_{S_u} dS(\boldsymbol{s})e^{ik\boldsymbol{s}\cdot\boldsymbol{r}}\mu(\boldsymbol{s},k), \tag{5}$$

where integration is over the unit sphere. The spherical harmonic expansion of the source-free region, using standard complex spherical harmonics Y_m^n corresponding to N3D harmonics Y_{mn}, is

$$\Psi(\boldsymbol{r},k)=\sum_m j_m(kr)\sum_n Y_m^n(\theta,\delta)A_m^n(k). \tag{6}$$

The terms $j_m(kr)$ are the spherical bessel functions. From Eqs. (6) and (5) valid plane wave coefficients can be found in terms of the spherical harmonic coefficients A_m^n (Duraiswami *et al.*, 2005)

$$\mu(\boldsymbol{s},k)=\sum_{m,n} i^{-n}A_m^n(k)Y_m^n(\boldsymbol{s}), \tag{7}$$

and in terms of the N3D convention

$$\mu(\boldsymbol{s},k)=\sum_{m,n} B_{mn}(k)Y_{mn}(\boldsymbol{s}). \tag{8}$$

The lack of a complex factor in Eq. (8) reflects the fact that in ambisonics, plane waves with zero phase at the center have real-valued encodings, allowing the identification to be made between microphone polar patterns and the N3D harmonics. From the linear supposition of plane waves, the binaural signals, $\Psi^L(k)$, $\Psi^R(k)$ are found by integrating the plane wave weights with HRTF responses, $H^L(k\boldsymbol{s})$, $H^R(k\boldsymbol{s})$, over the sphere

$$\Psi^L(k)=\int_{S_u} dS(\boldsymbol{s})\mu(\boldsymbol{s},k)H^L(k\boldsymbol{s}), \tag{9}$$

for the left side and similarly for the right. In practice the integral can be replaced by a quadrature sum, with very little loss of accuracy for a sufficient number of quadrature points, of order the number of spherical harmonics in Eq. (8) (Duraiswami *et al.*, 2005).

A high-order soundfield encoding can contain a variety of sources, both farfield and nearfield, so using the above method we are able to realize near sources binaurally with only knowledge of the the farfield HRTFs.

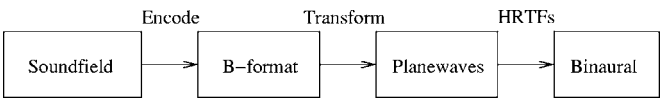

FIG. 2. Encoding a soundfield to binaural via a plane wave expansion.

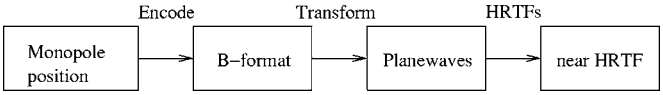

FIG. 3. Finding a nearfield HRTF from a monopole encoding.

Furthermore, by binaurally synthesizing a pure nearfield monopole of wave number k, the nearfield HRTF at that k, corresponding to the position of the monopole, is given immediately by $\Psi^L(k)$ in Eq. (9). Figure 3 summarizes this. The plane wave expansions of monopoles are investigated further in the next section.

Binaural rendering has the potential for greater realism than speaker array reproduction, because there is no constraint on the binaural signals delivered to the listener, whereas arrays must operate within a physical space and with the number of channels available, often in limited positions. On the other hand, binaural rendering has been hampered by the practical difficulties of ascertaining individual HRTFs and providing head tracking, both of which must be executed with precision. These problems are being addressed, and there are promising signs that they will be overcome. Therefore it is worthwhile to consider the process of binaural synthesis more carefully.

II. SCATTERING OF NEARFIELD SOURCES

A. Spherical expansion of a monopole

To study expansions of near sources we look in detail at the monopole. The important features are also true for general sources. A monopole source at noncentral position $\boldsymbol{r}'$ has the following expansion in $\boldsymbol{r}$, valid only for $r<r'$, Morse and Ingard (1968)

$$\frac{e^{-ik|r-r'|}}{|r-r'|}=ik\sum_{m=0}^{\infty} j_m(kr)h_m(kr')\sum_{n=-m}^{m} Y_{mn}(\theta',\delta')Y_{mn}(\theta,\delta), \tag{10}$$

where $j_m(kr')$ and $h_m(kr)$ are the spherical Bessel function of the first kind and spherical Hankel function of the second kind. The positive frequency convention is chosen which gives an outward moving wave for a time piece $e^{i\omega t}$. For the following discussion it is important to emphasize the region of validity for the expansion is a sphere extending as far as the source, as illustrated in Fig. 4. A slightly expanded restriction applies to more general sources that may extend over a region. In this case the valid region extends as far as the maximum radius that does not enclose any source.

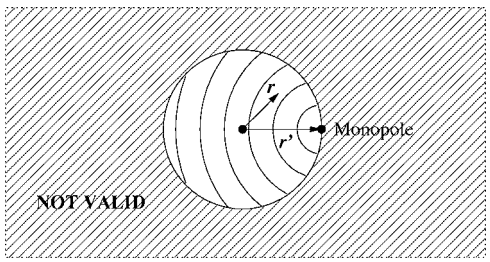

FIG. 4. Spherical expansion of a displaced monopole.

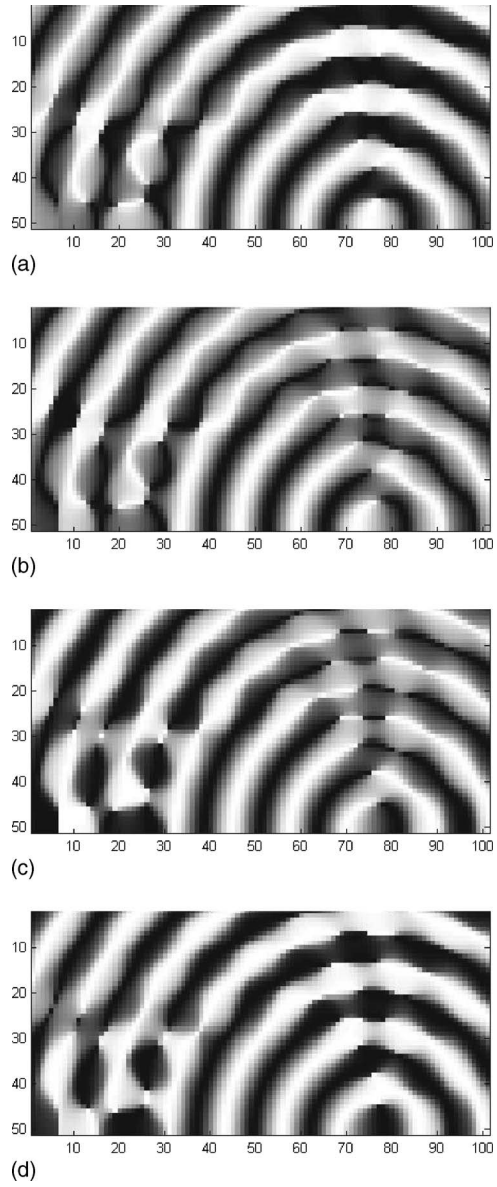

FIG. 5. Animation sequence, with $r'=2\lambda$, $m_{\max}=12$, $\omega t=0$, $\pi/4$, $2\pi/4$, $3\pi/4$.

A freefield expansion cannot be extended past the source completely because it must retain zero divergence everywhere. Figure 5 shows snapshots[1] taken at 1/4 cycle intervals of a half-plane cross section of an expansion with $r'=2\lambda$ and the maximum value of m, $m_{\max}=12$. The bessel functions $j_m(kr)$ are very close to zero for $kr<m$, so $m_{\max}\approx kr'$ is sufficient to synthesize accurately in the region $r<r'$. A detailed error analysis has been performed, (Duraiswami *et al.*, 2005). To aid visualization the plots are normalized $\mathrm{Re}(\Psi)/|\Psi|$. The monopole radiates outwards from the source in the valid region, while just outside the valid region, the field radiates inwards, satisfying zero divergence.

B. Plane wave decomposition

As described by Eq. (8), the spherical expansion can be re-expressed as a plane wave expansion. Figure 6 shows the expansion in Fig. 5 re-expressed with 196 plane waves, whose directions are distributed around the sphere on Fleige nodes, (Fliege and Maier, 1999). Fliege nodes are positioned for minimal quadrature error over the sphere. The valid region matches closely, the outer region has transformed and remains invalid.

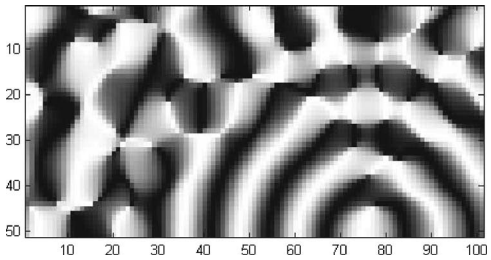

FIG. 6. Plane wave-from-spherical expansion of a monopole with $r'=2\lambda$, $m_{\max}=12{,}196$ Fliege nodes.

It appears at first that the angular bandwidth of the node set is sufficient if it matches that which is sufficient for a spherical harmonic expansion to radius r, $m_{\max}\approx kr$. This assumption was made in Duraiswami *et al.*, 2005. However, closer examination shows that a higher bandwidth for the node set, $m_{\max}\approx 2kr$, provides a nearer match.

C. Scattering validity of synthesized sources

A HRTF filter generates the signal at an ear resulting from the scattering of a plane wave by the listener. The phase of the plane wave is assumed to be zeroed to the center of the head. We can express the resultant field as a sum of the original unscattered field and the scattered component, $\Psi=\Psi_{\mathrm{in}}+\Psi_{\mathrm{scat}}$. The scattering can be formulated in terms of the Sommerfield radiation conditions, which state that Ψ_{scat} depends only on Ψ_{in} at the boundary of the scattering body. As we have seen, the ambisonic representation of a nearfield source is accurate only within a limited region, no matter how high the order of approximation. The derived plane wave decomposition can also only be accurate within the limited region. If part of the scattering body is outside this region, then Ψ_{in} is no longer correct on all of the scattering body. The binaural signals, found according to Eq. (9), are part of the resultant field Ψ, and so in general suffer loss of accuracy when the valid region does not enclose the scattering surface. Figure 7 illustrates this for a front view of a listener, where the source is close enough that everything from the shoulders downwards is excluded from the valid region. It has been shown previously that the torso plays a significant role in localization, (Algazi *et al.*, 2001).

The arrows in Fig. 7 show the flow of energy in the freefield expansion of the source, as illustrated previously in the animation sequence Fig. 5. In a field with a real source, both arrows would point away from the source. It is evident that the scattering in the shoulder region using the freefield will be quite different from the scattering with a source field, and cause differences in the resultant field at the ears.

D. Higher accuracy nearfield expansions

The above result is not too surprising in retrospect, because we should not expect to be able to construct the response from a source embedded in an arbitrarily complex

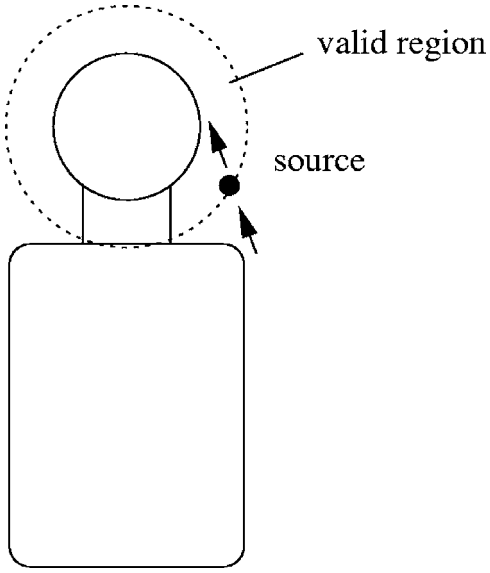

FIG. 7. Parts of scattering body outside the valid expansion region of a near source.

scattering geometry using only plane wave scattering responses. However, the question remains, how well can we do with plane wave HRTFs? Plane wave expansions have a useful property that spherical harmonic expansions lack, they can be translated and expressed about a different point simply by multiplying by phase factors: If a position relative to the new center is $\boldsymbol{r}'$, and the corresponding position relative to the old center is $\boldsymbol{r}$, then $\boldsymbol{r}=\boldsymbol{r}'+\boldsymbol{x}$, where $\boldsymbol{x}$ is the translation from the old to the new center. So $e^{i\boldsymbol{k}\cdot\boldsymbol{r}}=e^{i\boldsymbol{k}\cdot(\boldsymbol{x}+\boldsymbol{r}')}=e^{i\boldsymbol{k}\cdot\boldsymbol{x}}e^{i\boldsymbol{k}\cdot\boldsymbol{r}'}$. Therefore from Eq. (5) the expansion coefficients about the new center are $\mu'(\boldsymbol{k})=e^{i\boldsymbol{k}\cdot\boldsymbol{x}}\mu(\boldsymbol{k})$. Using this result, we can take the expansion for a source at a large radius, then shift the center so the source is at the required relative location. In the process the region of validity has been expanded, so a greater scattering body can be included, and the resulting binaural signals will be more accurate. Figure 8 illustrates this, showing a body entirely within the valid region for scattering, and with a source near to the head.

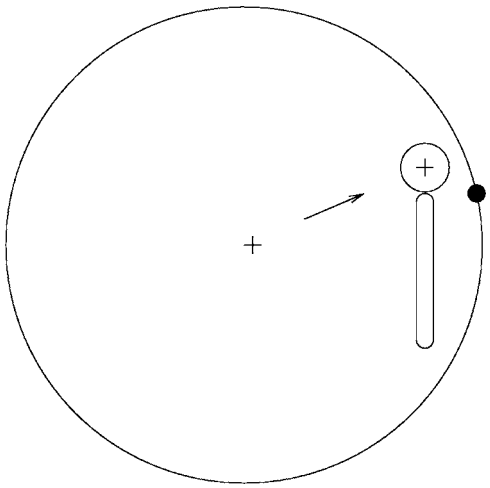

FIG. 8. Shifted plane wave expansion for full scattering with a near source.

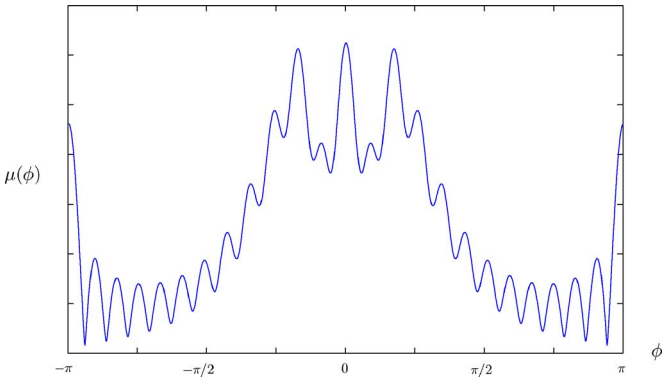

FIG. 9. Plane wave expansion coefficient for $r=2\lambda$, $m_{\max}=12$.

There is a cost to be paid for the improved rendering of the source, however. The new region of validity excludes other near source regions, which must be rendered separately with their own shifts, rather than from a single expansion. Also the new spherical expansion must be specified to higher order, because the region being scattered is at a greater radius in the freefield expansion. As a result of this the number of plane waves and hence HRTFs must be increased. The number of nodes required is $O(kr)$ where r is the radius of curvature of the region boundary.

The limit of the boundary of validity obtained by shifting is a plane through the source, so it would appear impossible, as conjectured earlier, to precisely generate binaural signals, using plane wave HRTFs, for sources in concave regions of the scattering body, such as under the chin. The fact that their exist a family of plane wave expansions about each point is a consequence of the presence of a source region, and the nonlocality of the plane wave. A source-free region on the whole plane, such as the harmonic expansion, has a unique plane wave expansion.

Figures 9 and 10 show two plots of the magnitude $|\mu(\phi)|$ of the plane wave decomposition of a freefield expansion of a monopole source at $r=2\lambda$, 4λ, respectively. The term ϕ is the spherical coordinate measuring the angle between the direction and the coordinate symmetry axis, with $\phi=0$ being the direction to the source. The decomposition of a source in a general position is just a rotation of $\mu(\phi)$. The greater detail seen in Fig. 10, compared with Fig. 9, reflects that it is sensitive to higher resolution HRTF data.

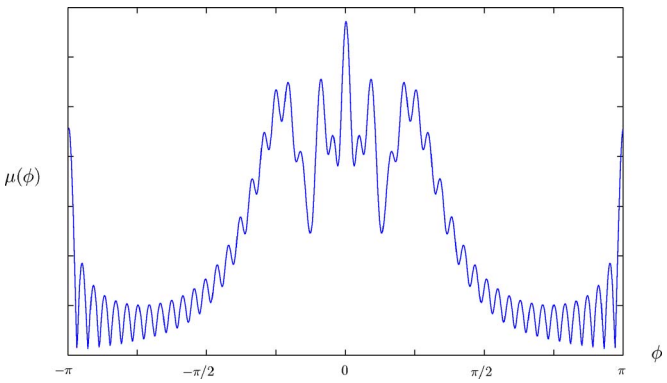

FIG. 10. Plane wave expansion coefficient for $r=4\lambda$, $m_{\max}=24$.

E. Complex sources

The argument developed above extends to complex localized sources, which can in general be written in terms of infinitesimal monopole clusters, or multipole expansions, and so subject to the same expansion region limitations. The valid region of a freefield expansion extends to the maximum radius that is free of any enclosed source.

The characterization of a general source by multipole expansion is a useful form of encoding. It can be transformed into a freefield expansion at the listener according to position and orientation, then into a plane wave expansion, and finally binaural signals. The details shall appear in a future publication.

F. Optimizing the expansion order

We go back now and consider the less special case of encoding sources that are not so close that the enlarge and shift process just described is neccessary to include relevant scattering. If the radius of the scattering region, centered on the head center, is r, then the order of expansion required is $m_{\max}=kr$. There is nothing to prevent us from moving the center of the expansion downwards to the center of the total scattering body, in order to minimize r, and therefore $m_{\max}$. The plane wave expansion must be phase shifted as before, prior to HRTF processing. A further refinement would be to separate the encoded signal into several banded signals. The lower frequencies would be diffracted by the body and require a larger scattering region, although lower sampling rates, while higher frequencies would require a smaller region at full rate.

G. Speaker arrays

In principle the limitation of ambisonic encodings applied to HRTFs is also relevant to array reproduction, except in this case the listener's body will scatter in real nonvalid regions, rather than via the HRTF filtering process. In practice near sources could only be produced for one listener, negating the main advantage of array systems, and even this would be difficult to achieve due to other limitations mentioned earlier. It is conceivable, however, that a specialist auditory display might be designed with a speaker array around a single listener, and then the consideration of scattering validity is relevant.

III. CONCLUSION

Ambisonic encoding provides a convenient soundfield representation, that can be rendered into binaural signals via plane wave decompositions in a precise way. However, there exists a surprising limitation for binaural rendering of near sources, namely errors are introduced due to incorrect accounting for scattering. It is possible to synthesize improved binaural signals from a given source by creating a more detailed plane wave expansion that allows for scattering objects in a larger region. To cover a general soundscape, we combine a main expansion covering most sources, with additional expansions for close source regions.

We have not presented any tests applying our methods with real HRTF data. This is clearly an important and complex task, which we hope to address. Working with high quality personalized HRTFs will be a key factor. With the ongoing refinement of virtual reality systems, the considerations presented here are expected to become increasingly relevant.

[1]A movie can be found at `www.cse.dmu.ac.uk/~dylan/NearfieldBinauralSynthesis/freefieldMonopole.avi` Last viewed 2/7/06.

Algazi, V., Avendano, C., and Duda, R. O. (**2001**). "Elevation localization and head-related transfer function analysis at low frequencies," J. Acoust. Soc. Am. **109**, 1110–1122.

Daniel, J. (**2000**). "Reprsentation de champs acoustiques, application la transmission et la reproduction de scnes sonores complexes dans un contexte multimdia," Ph.D. thesis, University of Paris 6, Paris, France.

Daniel, J. (**2003**). "Spatial sound encoding including near field effect," in *Proceedings of the AES 23nd International Conference*.

Duraiswami, R., Zotkin, D., Li, Z., Grassi, E., Gumerov, N., and Davis, L. (**2005**). "High order spatial audio capture and its binaural head-tracked playback over headphones with hrtf cues," in *Proceedings of the AES 119th Convention*, New York.

Fliege, J., and Maier, U. (**1999**). IMA J. Numer. Anal. **19.2**, 317–334.

Gerzon, M. (**1985**). J. Audio Eng. Soc. **33**, 859–871.

Gerzon, M. (**1992**). "General metatheory of auditory localization," in *Proceedings of the 92nd AES Convention*, Vienna.

Jot, J., and Wardle, S. (**1998**). "Approaches to binaural synthesis," in *Proceedings of the AES 105th Convention*, San Francisco.

McKeag, A., and McGrath, D. (**1996**). "Sound field format to binaural decoder with head tracking," in *Preprint 4302, AES Convention 6r*, Melbourne.

Menzies, D. (**2002**). "W-panning and o-format, tools for object spatialization," in *Proceedings of the AES 22nd International Conference*, Helsinki.

Morse, P., and Ingard, K. (**1968**). *Theoretical Acoustics* (Princeton University Press).