HOLOPHONIC SYNTHESIS OF MUSICAL SOURCES

Dylan Menzies
De Montfort University, Leicester, UK
dylan@dmu.ac.uk

Abstract

The composition of a soundfield containing complex nearfield sources has interesting musical applications. A method is presented here using freefield expansions about exterior points of a harmonic multipole field. A derivation and verification are included. Binaural rendering is the natural method for displaying such soundfields. Some possible musical applications are discussed.

INTRODUCTION

The spatial content of sound is an important part of how it is perceived. This has been reflected in the production of music through the ages, though the spatial acoustics of auditoria and the way sound is projected from musical sources. Through the development of the loudspeaker, and later digital signal processing, the possibilities for generating and controlling spatial sound has increased greatly. The traditions of electroacoustic and acousmatic music have explored these areas, and have sought to explicitly portray spatial qualities as musical attributes, both in recording and live performance. In mainstream recording, there is an increasing interest in spatial sound, as stereo is being superseded with 5.1 systems. Here we are concerned, in particular, with synthesizing complex virtual sources, with dynamic control over the location and orientation of the sources. It should be possible to capture and manipulate the experience of listening to a rich source such as a violin in close proximity, rather than working with the current static methods of recording. Similarly, purely synthetic sources could be created, which have spatial variety associated with traditional instruments.

Good results have been achieved for synthesizing distant sources, which reach the listener as plane waves. Distance perception can be simulated using distance filtering and reverberation balance. It is also possible in low-order Ambisonic systems, (Gerzon 1992; Gerzon 1985), to approximately synthesize a diffuse source at varying distance, (Menzies 1999; Menzies 2002), which can be useful in a creative setting. Soundfield synthesis of an object with non-uniform directivity has been considered in the farfield using spherical harmonic representation, (Menzies 1999; Menzies 2002). With the development of high-order acoustic field construction, the simulation of nearfield

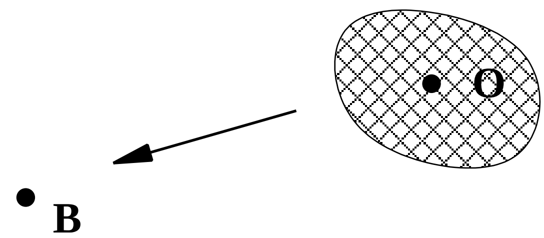

Figure 1: Overall scheme. O denotes the extended source object, and B the listener.

sources becomes feasible. This has been developed for a monopole source in the context of high-order Ambisonics by reconstructing a monopole field about the listener, (Daniel 2003). Using the wavefield approach, (Berkhout, Vries, and Vogel 1993; Daniel, Nicol, and Moreau 2003), the directional properties of objects have been encoded with filters that feed the speaker array directly, (Caulkins, Corteel, and Warusfel 2004; Warusfel 2004).

A localized source typically differs in two respects from simple monopole source. The sound radiates from a region of non-zero width, and the directivity of radiation is not uniform. Near to the object the soundfield will be reactive, like a monopole's, but possibly have a much more complex geometry. We should fully expect this added richness to be exploited by the auditory system for its information content, and so to have perceptual significance. Although this does not appear to have been studied in detail, informal listening provides strong evidence of spatial perceptual variety among complex objects. The study of directional objects using the wavefield approach also supports the hypothesis. For both practical and creative applications it would be desirable to find a way to accurately represent a complex source and encode it into Ambisonic B-format. The conversion from source encoding to Ambisonic encoding depends on the location and orientation desired of the source. From a single source encoding, that source can be rendered anywhere and in any orientation around the listener. Figure 1 illustrates this scheme. The advantage of Ambisonic modularity is apparent here, in that we seek a process that encodes into a format that is independent of the details of the rendering mechanism, whether it be a particular speaker array or headphones. The wavefield approach lacks this intermediate stage, as well as proving less accurate for given order in some studies (Daniel, Nicol, and Moreau 2003). Binaural rendering of high-order Ambisonics, over headphones, including the nearfield, has been considered, (Menzies and Al-Akaidi 2007b).

The article is organized as follows. First the source representation is discussed, followed by the main part, the development of a method to transform a source encoding, with knowledge of its position and orientation, into an Ambisonic encoding. Some simulations are provided for verification and illustration. Finally we consider how the approach can be adapted for the Ambisonic encoding of reverberation depending on source and listener positions and orientations. Owing to lack of space, this article has been shortened considerably. Interested readers are referred to a forthcoming article, (Menzies and Al-Akaidi 2007a).

SOURCE REPRESENTATION

We wish to use a representation which can encode any source to any desired accuracy, relates well to direct measurements of the field, and can be manipulated efficiently. The following possibilities suggest themselves. A source can be modelled with several monopoles. This would be appropriate if it actually has this structure, or because a rough and fast model is required. The source can be positioned and orientated using standard cartesian transformations. For more accuracy we can attempt to use many monopoles distributed over the source volume or surfaces. It is far from obvious how this would be done for a general source. Such a representation contains considerable redundancy since it describes the structure of the object as well as the sound produced.

The exterior harmonic expansion

Multipoles in their original form consist of infinitesimal arrangements of monopole sources. A multipole of sufficient order can represent a the field around a given extended object arbitrarily well. Although they are operated on by simple cartesian operations, their infinitesimal nature does not lend itself to direct numerical manipulation. Also the relationship of multipole parameters to the directionality of the field rapidly increase in complexity with order. Closely related is the *exterior expansion* for the wave equation. This has basis functions in the frequency domain using spherical coordinates, $h_m(kr)Y_{mn}(\theta,\delta)$, where $h_m(kr)$ are the spherical hankel functions of the second kind, (Morse and Ingard 1968). m is the multipole order of each function, and $k = 2\pi/\lambda$ is the wavenumber. The type of hankel function chosen gives an outward moving wave when associated with a positive frequency time piece $e^{i\omega t}$, the same convention used in (Daniel 2003).

An infinitesimally defined multipole of order m can always be expressed exactly using an exterior expansion with terms up to order m. For this reason an exterior expansion is alternatively called an *exterior multipole expansion* or just a *multipole*, (Gumerov and Duraiswami 2005). Another term used is *singular expansion*, since the center of the expansion has a singularity. The exterior expansion relates closely to the non-uniform directivity of a source, as discussed below, and our principal goal shall be to manipulate it to provide an Ambisonic source encoding. By multipole we shall mean an exterior expansion, unless otherwise stated.

For convenience we define coefficients, $O_{mn}(k)$, by a general exterior expansion,

$$p(\boldsymbol{r},k) = k\sum_m i^{-m-1}h_m(kr)\sum_n Y_{mn}(\theta,\delta)O_{mn}(k)\ , \tag{1}$$

so that in the farfield where $h_m(kr)$ tends to $i^{m+1}e^{-ikr}/kr$, the field becomes

$$p_{far} = \frac{e^{-ikr}}{r} \sum_{m,n} Y_{mn}(\theta, \delta) O_{mn}(k) \ . \tag{2}$$

The $O_{mn}(k)$ coefficients then directly express the non-uniform directivity in this regime, where locally the field tends to an outward moving plane wave. The signals $O_{mn}(k)$ coincide with the *O-format* encoding used previously for Ambisonic synthesis, (Menzies 1999; Menzies 2002). The same name will be used here for the more general case described by (1). This is just a convenient convention, in the same sense as B-format is defined. Nothing essentially new is added.

$O_{mn}(k)$ can be readily calculated from measurements of the field on a sphere at any radius r outside the source region. Applying an integral over the sphere, $\int d\Omega \ Y_{mn}(\theta, \delta)$ to (1) gives

$$O_{mn}(k) = \frac{i^{m+1} \int d\Omega \ Y_{mn}(\theta, \delta) p(\boldsymbol{r}, k)}{4\pi k h_m(kr)} \ . \tag{3}$$

Source approximation order and error

We consider now the order to which a source is approximated, m_{max}. We wish to minimize this subject to reconstruction error constraints. A source can be arbitrarily small and still have power up to any multipole order, for example using the explicit definition of infinitesimal multipoles. However this is unusual in a real acoustic source because opposed component sources are not usually found very close together. Here we just have room to give a general result for typical extended sources. For object radius r, the maximum order required in the expansion, $m_{max} \approx kr$.

AMBISONIC ENCODING OF MULTIPOLES

Freefield expansion

High-order Ambisonics is founded on the *interior expansion* that we shall also call the *freefield expansion* here, to emphasize that it is used to describe a sourceless region around the listener. Eq. (4) is the version of the expansion using N3D harmonics, $Y_{mn}(\theta, \delta)$, and defines the B-format coefficients, $B_{mn}(k)$, (Daniel 2003). The expansion converges quickly on any source-free field, up to a given radius r. The typical order required to achieve $\approx 1\%$ error for a regular freefield, such as a planewave, is $m_{max} \approx kr$, (Gumerov and Duraiswami 2005; Ward and Abhayapala 2001).

$$p(\boldsymbol{r}, k) = \sum_m i^m j_m(kr) \sum_n Y_{mn}(\theta, \delta) B_{mn}(k) \tag{4}$$

A freefield expansion is by definition a sourceless field. This means that it can only be extended in radius as far as the nearest source, which however does not prevent us recreating the freefield around a listener due to a nearby source.

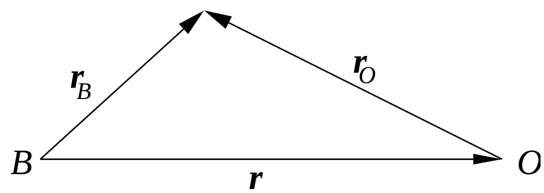

Figure 2: Vector notation

Multipole to freefield coefficient transformation

The main task in this section is to find $B_{mn}(k)$ in the presence of a multipole described by $O_{mn}(k)$ at a given position. It would be desirable to find a generalized closed form expression, as for the monopole case in (Daniel 2003). However, it is not very apparent how this could be done or even if it would be the most practical method of calculation, so instead a more pragmatic approach is adopted yielding eventually a manageable integral expression. To begin (4) and (1) are equated. The notation is modified according to Figure 2,

$$\sum_m i^m j_m(kr_B) \sum_n Y_{mn}(\theta_B, \delta_B) B_{mn}(k) = k \sum_m i^{-m-1} h_m(kr_O) \sum_n Y_{mn}(\theta_O, \delta_O) O_{mn}(k) \tag{5}$$

To isolate $B_{mn}(k)$ the operator $\int d\Omega_B \; Y_{m'n'}(\theta_B, \delta_B)$ is applied, with r_B a freely chosen constant, and θ_O, δ_O and r_O are functions of the vector $\boldsymbol{r}_B$, yielding

$$4\pi i^{m'} j_{m'}(kr_B) B_{m'n'}(k) = k \sum_m i^{-m-1} \sum_n O_{mn}(k) \int d\Omega_B \; Y_{m'n'}(\theta_B, \delta_B) Y_{mn}(\theta_O, \delta_O) h_m(kr_O). \tag{6}$$

After some manipulations, described in (Menzies and Al-Akaidi 2007a), the following expression for $B_{mn}(k)$ is found,

$$B_{mn}(k) = \sum_{n'} R_{mnn'}(\theta, \phi) \sum_{m'} \frac{1}{r} M_{mn'm'}(kr) \sum_{n''} R_{m'n'n''}(\theta, -\phi) \; O_{m'n''}(k) \tag{7}$$

where the new 3-index coefficient matrix is defined,

$$M_{mnm'}(k) = \frac{k i^{-m-m'-1}}{2 j_m(k_B)} \int_{-1}^{+1} ds_B \hat{P}_{mn}(s_B) \hat{P}_{m'n}(s_O) h_{m'}(k_O), \tag{8}$$

and,

$$\hat{P}_{mn}(\sin\delta) = \sqrt{(2m+1)\frac{(m-|n|)!}{(m+|n|)!}} \; P_{m|n|}(\sin\delta) \quad . \tag{9}$$

$R_{mnn'}$ denotes a harmonic rotation, and (θ, ϕ) describes the location of the source. $P_{mn}(x)$ is the associated Legendre polynomial. $k_B = \alpha k$, $k_O =$

$k\sqrt{1+\alpha^2-2\alpha s_B}$, $s_O = r(\alpha s_B - 1)/r_O$ and $s_B = \sin\delta_B$. Clearly this is defined only for $n < m$ and $n < m'$, so for a given source the number of filters increases only linearly with B-format order required. The filter coefficients are given in terms of one parameter, k. The actual filter acting in (8) is scaled in frequency by the radius r and there is a distance factor $1/r$. Choosing $\alpha = m/k$ gives good numerical performance.

Validation and properties

To provide an immediate confidence test that the derived formulas are correct, a random test 5th order multipole was constructed, shown in Figure 3, and compared with the 13th order freefield expansion calculated using the matrix (7), shown in Figure 4. The error contours in Figure 4 at 10% and 1% levels are for deviations from the original multipole shown in Figure 3. The region of agreement extends as far as the center of the original multipole, as expected, and indicates that the calculations described in this section are correct. Further tests, show that (7) agrees with the monopole case previously considered. The general picture for higher orders is that with e^{-ikr}/r factored out, the response is always minimum phase. For small k the order of the filter becomes $m + m'$, while for large k it is n. The location of the transitional region increases linearly with $m+m'$ from $k \approx 2$ for $m+m' = 1$. For higher orders, the transitional region can be more complex. The transfer functions grow large as k tends to zero. It can be shown that the functions can be limited without significantly affecting accuracy in the regions of interest.

With rotations included, the number of filters required to synthesize a given source is only linear in the maximum B-format order, m_{max}, for $m_{max} > m'$, owing to zeros in $M_{mn'm'}(kr)$, although the complexity of the filters themselves increases with m. As we saw earlier, $m_{max} \approx rk$, where r is the radius of the 3-dimensional region constructed accurately. This demonstrates the inherent efficiency of the method, despite the precision and complexity of reconstruction.

Reverberation encoding and transformation

Related results have also been found for the generation and transformation of reverberation, (Menzies and Al-Akaidi 2007a). This provides an efficient means for a listener to experience a region of high quality reverberant sound field, while being freely able to move and orientate their heads in that region.

CONCLUSION

A method has been presented for encoding a general acoustic source, and transcoding it to a high-order Ambisonic signal, dependent on source orientation, and position relative to the listener. The method also lends itself to the direct measurement of real sources using an array of surrounding microphones, as well as more indirect synthetic processes. As well as a recording and mixing tool, it can be seen as a compositional tool, and in the context

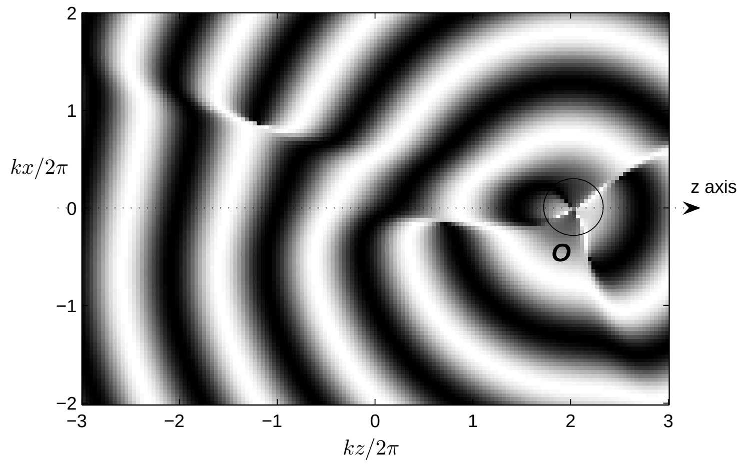

Figure 3: Cross-section of a field plot for a 5th order multipole, center at **O**. The cross-section is $\theta = 0$. x, z are cartesian coordinates in length units.

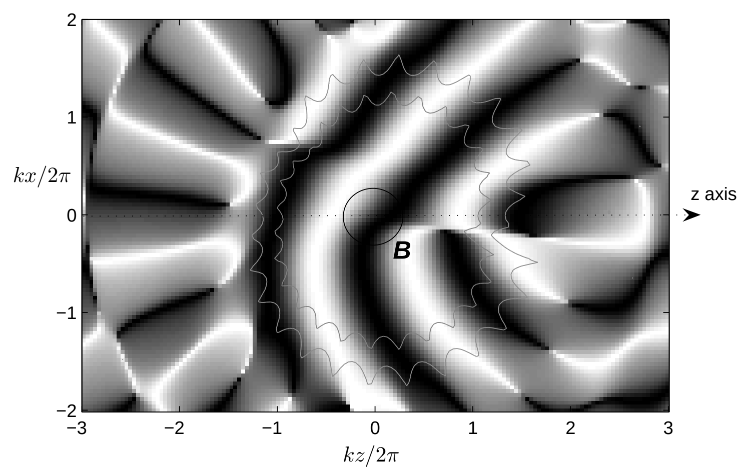

Figure 4: Cross-section of a field plot for a 13th order freefield expansion, center at **B**, of a multipole, center **O**. Error contours are shown at the 1% and 10% levels. The cross-section is $\theta = 0$. x, z are cartesian coordinates in length units.

of progressive electronic music, a performance instrument. The approach is considerably more elaborate and costly than plane wave or monopole synthesis, however it is expected that in the context of complex sources displayed with a high-quality rendering system, the efforts are worthwhile. Binaural headphone reproduction is particularly attractive, because the encoding only needs to be of sufficient order for a single listener rather than a listening area, so reducing computational costs. For instance for radius 0.2 m, up to 1500 Hz, the required order, $m = rk \approx 6$. In a speaker rendering environment the valid listening region is necessarily fixed to accommodate multiple listeners. This places constraints on how nearfield sources can be arranged relative to the listener, so for example it is impossible for a listener to experience near sources directly on the left and the right sides while also having a large listening area that can hold multiple listeners. Binaural reproduction does not suffer this constraint, and so is the more natural method for nearfield rendering. A possible exception would be a small speaker array designed for one person. In the future we hope to investigate realizations of these methods.

REFERENCES

Berkhout, A., D. Vries, and P. Vogel (1993). Acoustic control by wave field synthesis. *J. Acoust. Soc. Am. 93*, 2764–2778.

Caulkins, T., E. Corteel, and O. Warusfel (2004). Synthesizing realistic sound sources in wfs installations. In *Proc. DAFX04*.

Daniel, J. (2003). Spatial sound encoding including near field effect. In *Proc. AES 23nd International Conference*.

Daniel, J., R. Nicol, and S. Moreau (2003). Further investigations of high order ambisonics and wavefield synthesis for holophonic sound imaging.

Gerzon, M. (1985). Ambisonics in multichannel broadcasting and video. *Journal of the Audio Engineering Society 33*, 859–871.

Gerzon, M. (1992). General metatheory of auditory localisation. In *92nd AES Convention Preprints*.

Gumerov, N. A. and R. Duraiswami (2005). *Fast multipole methods for the Helmholtz equation in three dimensions*. Elsevier Science.

Menzies, D. (1999). *New Performance Instruments for Electroacoustic Music*. Ph. D. thesis, University of York, UK.

Menzies, D. (2002). W-panning and O-format, tools for object spatialisation. In *Proc. AES 22nd International Conference*.

Menzies, D. and M. Al-Akaidi (2007a). Ambisonic synthesis of complex sources. *Journal of the Audio Engineering Society*, To appear.

Menzies, D. and M. Al-Akaidi (2007b, March). Nearfield binaural synthesis. *J. Acoust. Soc. Am. 121(3):1559-63.*

Morse, P. and K. Ingard (1968). *Theoretical Acoustics*. McGraw-Hill.

Ward, D. B. and T. D. Abhayapala (2001, September). Reproduction of a plane-wave sound field using an array of loudspeakers. *IEEE Transactions on Speech and Audio Processing 9(6)*, 697.

Warusfel, O. (2004). Reproduction of sound source directivity for future audio applications. In *Proc. International Congress on Acoustics*.