Audio Engineering Society

Convention Paper 8878

Presented at the 134th Convention
2013 May 4–7 Rome, Italy

Binaural Ambisonic Decoding with Enhanced Lateral Localization

Tim Collins[1]

[1] *School of Electronic, Electrical and Computer Engineering, University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK*

Correspondence should be addressed to Tim Collins (T.Collins@bham.ac.uk)

ABSTRACT

When rendering an ambisonic recording a uniform speaker array is often preferred with the number of speakers chosen to suit the ambisonic order. Using this arrangement, localization in the lateral regions can be poor but can be improved by increasing the number of speakers. However, in practice this can lead to undesirable spectral impairment. In this paper, a time-domain analysis of the ambisonic decoding problem is presented that highlights how a non-uniform speaker distribution can be used to improve localization without incurring perceptual spectral impairment. This is especially relevant to binaural decoders, where the locations of the virtual speakers are fixed with respect to the head, meaning that the interaction between speakers can be reliably predicted.

1. INTRODUCTION

1.1. Ambisonics

Ambisonics is a spatial audio encoding method used to record an approximation of a two- or three-dimensional sound field [1]. A distinctive feature of ambisonic encoding is that there is no information stored in, or implied by, the recording that dictates the geometry of the speaker array used to render it. Various speaker geometries are possible and specific decoders are required in each case. The accuracy and stability of the reproduction depends on factors such as the order of the ambisonic recording [2], the speaker placement and the number of speakers used [3][4].

1.2. Lateral Localization

A phenomenon common to many spatial audio systems is that the localization accuracy of sound images degrades for sources located in the lateral regions to the far left or right of the listener [5]. Am-

bisonic decoders are generally designed in order to optimize the velocity and energy vectors when an encoded plane wave is reproduced [6]. Optimization of the velocity vector and energy vector will ensure that the interaural time delay (ITD) and interaural level difference (ILD) are correct, however, it is likely that the pattern of pinna-related notches in the high-frequency spectrum will not be identical to that of a genuine single plane wave source [7]. For most source positions this is not problematic because the pinna response is only used to disambiguate the source location around the cone of confusion formed by the set of source angles that produce identical ITDs and ILDs (as described further in section 3). If the pinna response does not match any of the angles around the cone, the human auditory system assumes the most likely estimated source position around the cone with the closest match. For most source locations, this is enough information for the auditory system to make a sufficiently accurate approximation. In the lateral regions, however, the relative time delays that form the pinna response change rapidly with angle and the cone of confusion narrows. These factors combine to make localization of phantom images hardest in this region and can lead to smearing of the perceived image location [5]. This confusion is especially problematic with binaurally rendered sound sources because head movement cannot be used to assist with localization as it would in a natural environment. The causes of poor localization in the lateral regions are explored further in section 3 of this paper.

1.3. Speaker Distribution

An obvious approach to the issue of poor localization is to increase the number of speakers in the array (virtual speakers in the case of a binaural decoder). This reduces the maximum possible angle between an image location and the nearest speaker, thereby reducing the disparity between the ideal and reproduced pinna response patterns. Unfortunately, when the number of speakers in an ambisonic system is increased beyond the minimum requirement, spectral impairment in the high frequency region results [2]. The impact of this effect is described in section 3.3.

Spectral impairment effects can be modeled for plane waves reproduced by large numbers of speakers using either time- or frequency-domain analysis. Frequency-domain analyses are most often used because they directly yield the spectral response of the system [2]. However, the time-domain analysis, presented in section 4, is also revealing because it gives insight into how the problem can be ameliorated.

Analyses of the time-domain responses from various source angles show that the maximum speaker density allowable before audible spectral impairment results is not a constant for a specified ambisonic order, but varies according to the source position. Section 5 of this paper shows how this phenomenon can be exploited by designing non-uniformly distributed arrays with greater speaker density in the lateral regions. The analysis assumes that the head position is fixed and known. Although this cannot be generally assumed, it is the case for binaural decoders where the problem of lateral localization is most severe.

2. AMBISONIC ENCODING AND DECODING

2.1. Ambisonic Encoding

Ambisonic encoding is used to represent a two- or three-dimensional acoustic pressure field as a function of cylindrical or spherical harmonic components, respectively. Taking the horizontal-only (two-dimensional) case, the field can be expressed by a two-dimensional spectrum, $S(k, \theta)$, which gives the complex amplitude of the plane-wave component of the field arriving from angle, θ, as a function of wavenumber, k. This can be expressed as a sum of cylindrical harmonic components as [8]:

$$S(k,\theta) = \sum_{m=-\infty}^{\infty} q_m(k) e^{jm\theta} \tag{1}$$

where $q_m(k)$ are the spectra of the harmonic coefficients. In a practical ambisonic system, the number of channels is limited by the ambisonic order of the system, M, giving the truncated expression:

$$S(k,\theta) = \sum_{m=-M}^{M} q_m(k) e^{jm\theta} \tag{2}$$

For a plane wave with an angle of incidence, θ_0, and amplitude, A, the q_m coefficients are independent of wave number and can be expressed as [8]:

$$q_m = A e^{-jm\theta_0} \tag{3}$$

NB. The more familiar B-format channels can be straightforwardly derived from the q_m coefficients [2]. For example, for a first order system:

$$\begin{aligned} W &= \frac{q_0}{\sqrt{2}} \\ X &= \frac{q_1 + q_{-1}}{2} \\ Y &= \frac{q_1 - q_{-1}}{2j} \end{aligned} \tag{4}$$

Although B-format is a more commonly used convention, the notation used in equation 2 will be used throughout this paper because it is a more compact form that makes the analysis in section 4 more straightforward.

2.2. Ambisonic Decoding

Unlike most spatial audio formats, ambisonic encoding does not place any requirements on the geometry of the speaker array used to render it. This means that a specific decoder is required for each system to appropriately decode the ambisonic channels into speaker feeds. The decoding process is not simply a matter of reproducing the harmonic components exactly. For finite speaker arrays, this is only possible at the origin and it will be seen in section 3.3 that unwanted spectral impairment can result for offset listening locations if this approach is followed.

The approach generally adopted when designing ambisonic decoders is to attempt to produce the best possible perceptual impression of the encoded soundfield. The criteria used to assess the quality of a decoder concern the uniformity of the system response with angle and the accuracy with which plane waves are represented across the audible band [6]. Given a required speaker geometry, various algorithms have been developed to optimize these criteria for standard surround-sound geometries [9] and for more general two- and three-dimensional arrays [6].

2.3. Binaural Ambisonic Decoding

Binaural reproduction requires the decoding of the ambisonic channels to just two headphone output channels: one feeding directly into each ear. Conceptually, the procedure has two stages. First, the ambisonic signal is decoded in the conventional manner to produce a set of feeds to a virtual speaker array.

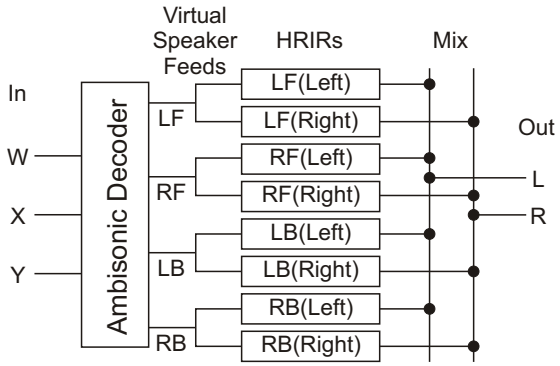

Fig. 1: Block diagram of a simple first-order 2D binaural decoder. Four virtual speakers are used in a square array consisting of: Left-Front (LF), Right-Front (RF), Left-Back (LB) and Right-Back (RB). WXY are the B-format ambisonic input channels and LR are the left and right headphone feeds.

Each speaker feed is then convolved with the Head Related Impulse Response (HRIR) corresponding to the position of the virtual speaker which, ideally, will be customized to match the HRIR that the listener would experience in a real listening situation. Finally, the stereo outputs of each HRIR-convolved channel are mixed together to form the single stereo feed to the headphones. Figure 1 illustrates a simple example of this process. In this example, a first-order 2D ambisonic input is decoded to feed four virtual speakers, each of which is modeled by convolution with an HRIR pair [10].

In practice, the binaural decoding processing can be simplified by noting the linear nature of the decoding matrix and HRIR processing. The ambisonic decoder and HRIR convolution can be combined for each ambisonic channel into a single pair of linear filters. In the case shown in figure 1, this would reduce the number of convolution filters required from eight to six (two times the number of ambisonic input channels). In general, this means that the complexity of the decoder is only dependent on the ambisonic order and not on the number of virtual speakers [10].

When compared with a conventional ambisonic decoder, a binaural decoder has the advantage that the listener position is known exactly and can be located

precisely in the center of the virtual array's 'sweet spot'. However, it is disadvantaged by the lack of head-movement and environmental cues which aid localization. These cues will be further discussed in section 3.

A further potential advantage of binaural decoding is that the geometry of the virtual speaker array is not restricted by the usual limitations of room geometry, aesthetic considerations or, simply, cost. Instead, it is possible to create entirely arbitrary virtual speaker geometries designed to optimize the perceived soundfield.

3. LOCALIZATION OF SOUND SOURCES

3.1. Localization Cues

The human auditory system uses various cues to localize a sound source:

- Interaural Level Difference (ILD): The amplitude difference of the sound heard by either ear caused by head shadowing effects.
- Interaural Time Delay (ITD): The relative time delay caused by the difference in propagation path lengths to the two ears.
- Pinna filtering: The shape of the outer ear introduces several reflections whose amplitude and relative delays vary with angle of incidence.
- Head movement: Correlation between head movement and the perceived location of a sound source.
- Environmental cues: The early reflections heard in a typical reverberant room can also assist localization in terms of angle of incidence and distance.
- Visual cues: If a listener can see an obvious source of a sound, this cue can override the perceptual effect of all others.

In most scenarios, ILD and ITD are the most significant localization cues. ITD is particularly important at low frequencies where head-shadowing effects are less pronounced and the ILD cannot be reliably used. At high frequencies, however, when the wavelength of the sound source is less than the head diameter, the ITD becomes ambiguous but the ILD is more significant because of increased head-shadowing attenuation. To achieve good localization performance across the entire audible band, decoders generally attempt to optimize both the perceived ILD and ITD, although not, necessarily, with the same emphasis at all frequencies [6].

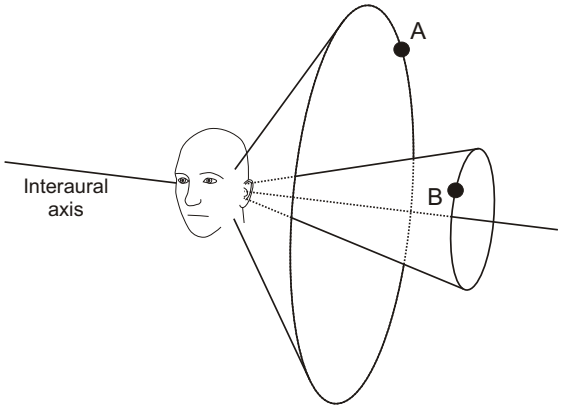

Fig. 2: Cones of confusion for two example source locations, 'A' and 'B'.

Although the primary measures used for localization are the ILD and ITD, these are not sufficient, alone, to uniquely position a sound source. Assuming a simplified spherical head model, for any source position there exists a cone of potential source angles for which the ILD and ITD are identical. This is often referred to as the "cone of confusion". This cone is symmetrical about the interaural axis and has its apex at the center of the head. Figure 2 shows two such cones for two example source locations labeled 'A' and 'B'.

In order to uniquely locate a sound, additional auditory cues must be used such as pinna filtering, head movement, environmental cues (e.g. early reflections) or visual cues. In many systems, environmental or visual cues may not be present and it is not always possible to make use of head movement (e.g. for short duration sounds or rapidly moving sources). Pinna effects are, however, almost always of use except with sounds that are lacking in high frequency content. In particular, when using binau-

ral reproduction systems it is often only ILD, ITD and pinna filtering that are available.

3.2. Difficulties in the Lateral Regions

The pinna response varies as a function of the azimuth and elevation of the sound source relative to the interaural axis. In the examples illustrated in figure 2, sound source 'A' has a relatively broad cone of confusion. As a consequence, although all angles on the cone share the same ILD and ITD, there is a significant variation in pinna response allowing good localization. By contrast, sound source 'B' has a more acute cone of confusion. Pinna response varies much less around the smaller solid angle described by the cone and localization is correspondingly much less precise.

Lateral localization problems with surround sound systems have been identified for a long time [5]. Taking a simple quadraphonic example scenario as illustrated in figure 3, the system is attempting to produce a phantom image using the two speakers RF and RB. For a well designed panning law, the image produced will have the correct ILD and, ideally, also the correct ITD placing it on the correct cone of confusion. The two speakers will have very different pinna responses with notches at different characteristic frequencies. Simply mixing them will not produce the actual pinna response of the phantom image angle. In this situation, the human auditory system will normally infer the most likely source position based on the best match with the received pinna response. Around the lateral region, however, pinna notch locations vary rapidly with angle and the phantom image will have a very different response to the ideal one leading to poor localization and an ambiguous, relatively diffuse phantom image.

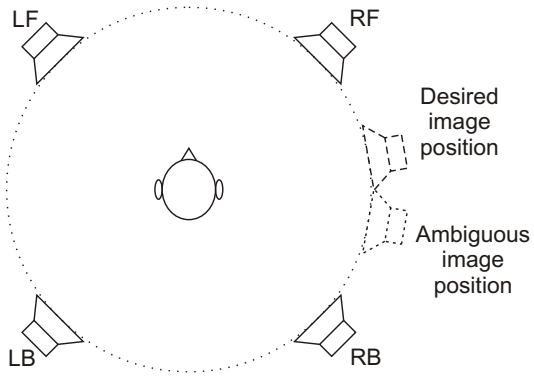

Fig. 3: Lateral Localization in surround sound.

3.3. Spectral Impairment

The obvious way to alleviate the problems described in section 3.2 is to increase the number of speakers used. The closer the phantom locations are to an actual speaker position, the less the error in the pinna response will be. Unfortunately, it is not possible to simply increase the speaker density arbitrarily. For a 2D ambisonic system of order, M, as soon as the number of speakers used exceeds $(2M + 1)$, spectral impairment results in the form of low-pass filtering [2]. This filtering effect varies with source angle and also features deep notches in the frequency response making equalization compensation only partially effective. A more effective approach would be to suppress or eliminate the cause of the filtering rather than to attempt to correct for it afterwards.

In order to increase speaker density in a way that minimizes spectral impairment problems, a non-uniform speaker distribution is proposed. The rationale behind this proposal stems from analysis of the responses of ambisonic decoders. Although frequency-domain analysis has been used to identify spectral impairment in the past [2], time-domain analysis offers an alternative perspective and provides insights into the method that can be used to minimize its impact.

4. TIME-DOMAIN ANALYSIS

4.1. Theoretical Analysis

In order to analyze the time-domain impulse response of a practical ambisonic decoder, the approach will be to first examine the case of an infinitely dense circular array of speakers. This is the ideal scenario if the listener is exactly located at the center of the array [1]. It is also the limiting case for finite uniform speaker arrays when the speaker density is progressively increased. Following this analysis, the case of finite uniform speaker density will be considered and a proposed non-uniform distribution introduced.

In order to evaluate the possible impairment effects, the time-domain impulse response of the system will be calculated. This is formed assuming an ideal impulse function is transmitted as a plane wave with an angle of incidence, θ_0, at such a time that it will arrive at the center of the array (the origin) at $t = 0$. Starting with the expression for the acoustic pressure field from equation 2, the time-domain pressure field observed at the origin can be expressed as a function of time and angle:

$$p(t, \theta) = \delta(t) \sum_{m=-M}^{M} q_m e^{jm\theta} \tag{5}$$

In practice, it is impossible for a human listener to have both ears located exactly at the origin because of their head diameter. The actual listening positions will be a distance, r, from the center of the array. The simplest scenario to analyze involves considering just one of the ears at a time and assumes that the listening position for that ear lies on the x-axis at coordinates $(r, 0)$. This geometry is illustrated in Figure 4.

If an impulse is transmitted from the speaker array, the plane-wave arrivals from all angles converge at the origin at time, $t = 0$. For any listening position with a non-zero value of r, however, this will not be the case. At an offset listening position, the impulse will be smeared in time between $t = \pm r/c$. The arrival time can be calculated as a function of angle:

$$t_\theta = -\frac{r}{c} \cos\theta \tag{6}$$

An implication of equation 6 is that at any time, t, between $\pm r/c$, the sound pressure received will be the sum of the pressures arriving from the two angles: $\pm \cos^{-1}(-tc/r)$. To calculate the observed pressure as a function of time, we first consider a small interval of time, ∂t, centered around a time, t. The observed pressure during this interval will originate from a pair of arcs, each with an extent of $\partial\theta$ and centered at $\theta = \pm \cos^{-1}(-tc/r)$. From each arc, the impulse function will be scaled by a factor of $\partial\theta/2\pi$ and the average pressure during the duration, ∂t, can, therefore, be calculated:

$$p(r, t) = \frac{1}{\partial t} \frac{\partial\theta}{2\pi} \text{rect}\left(\frac{ct}{2r}\right) \sum_{m=-M}^{M} \left[q_m e^{jm\theta} + q_m e^{-jm\theta}\right] \tag{7}$$

The rect() function constrains the response to be zero for $|t| > r/c$. Combining the complex exponentials simplifies the equation to:

$$p(r, t) = \frac{1}{2\pi} \frac{\partial\theta}{\partial t} \text{rect}\left(\frac{ct}{2r}\right) \sum_{m=-M}^{M} 2q_m \cos m\theta \tag{8}$$

In the limit as ∂t tends to zero, equation 8 simplifies to:

$$p(r, t) = \frac{c}{\pi r \sin\theta} \text{rect}\left(\frac{ct}{2r}\right) \sum_{m=-M}^{M} q_m \cos m\theta \tag{9}$$

Noting that

$$\cos(m\theta) = T_m(\cos\theta) \tag{10}$$

where T_m is the m-th order Chebyshev polynomial [11], and substituting from equation 6 gives the result:

$$p(r,t) = \frac{c}{\pi r \sqrt{1 - (ct/r)^2}} \text{rect}\left(\frac{ct}{2r}\right) \sum_{m=-M}^{M} q_m T_{|m|}\left(-\frac{ct}{r}\right) \tag{11}$$

Finally, considering a plane-wave arriving from an

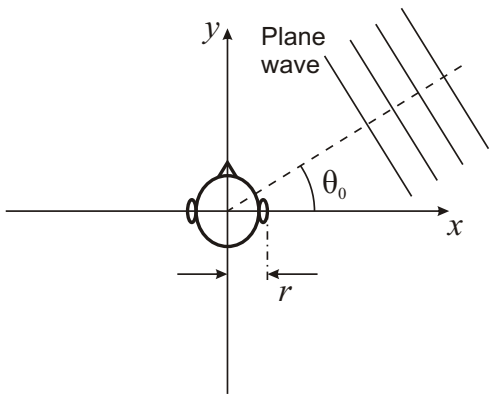

Fig. 4: Geometry of the practical listening position used in the time-domain analysis.

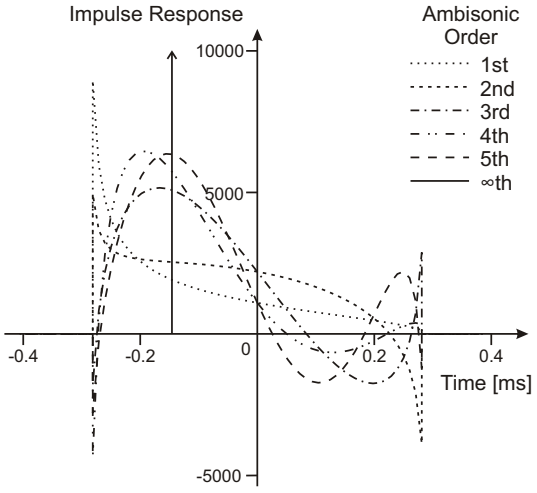

Fig. 5: Impulse responses for a plane wave source at an angle of -30° with different ambisonic encoding orders ($r = 0.1$ m).

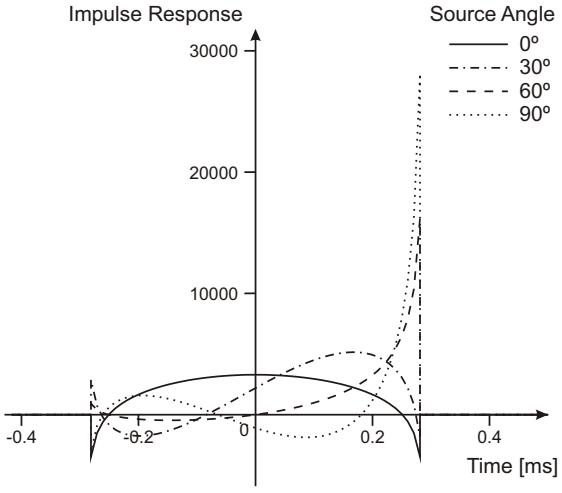

Fig. 6: Impulse responses for plane wave sources at different angles using 3rd order ambisonic encoding ($r = 0.1$ m). The responses for negative angles are not shown, for reasons of clarity, but can be found by reflection about the $t = 0$ axis.

angle, θ_0 and substituting from equation 3:

$$p(r,t) = \frac{c}{\pi r\sqrt{1-(ct/r)^2}}\text{rect}\left(\frac{ct}{2r}\right)\sum_{m=-M}^{M} e^{-jm\theta_0} T_{|m|}\left(-\frac{ct}{r}\right) \quad (12)$$

Because of the isotropic nature of ambisonic encoding, equation 12 can also be generalized as the result for any receiver location at a distance, r, from the origin where θ_0 is now the angle between the receiver location vector and the source direction.

4.2. Example Impulse Responses

Equation 12 can be used to predict the impulse response of an ambisonic decoder using an infinitely dense circular array of speakers as a function of ambisonic order, M, and of source angle, θ_0. Figure 5 shows the effect of increasing the ambisonic order for a fixed source angle of -30° and a listener location 0.1 meters to the left of the origin. Ambisonic orders from first to fifth are illustrated as well as the ideal case of an infinite order system where the response is simply a time-shifted version of the transmitted impulse.

It can be seen from figure 5 that all of the finite-order impulse responses indicate a degree of time smearing limited between $t = \pm r/c$. This will cause low-pass filtering effects above around $c/2r$ Hz. Although the response does clearly tend towards the infinite order case with increasing order, there is not a significant reduction in the time spread for currently practicable system orders.

The distribution of energy within the range $t = \pm r/c$ is not uniform and will also vary with source angle. Figure 6 shows the variation in impulse response for a 3rd order system. At zero degrees (i.e. directly in front of the listener), the response is spread in a broad lobe extending across most of the limiting interval, $\pm r/c$. However, as the source angle approaches the lateral regions ($\geqslant 60°$ in figure 6) the distribution becomes markedly concentrated around a much shorter interval. The result is a reduction of the effective duration of the impulse response which leads to an expansion of the usable bandwidth in the frequency-domain.

4.3. Finite Number of Speakers

A practical system will, of course, have only a finite number of speakers. For uniformly distributed ar-

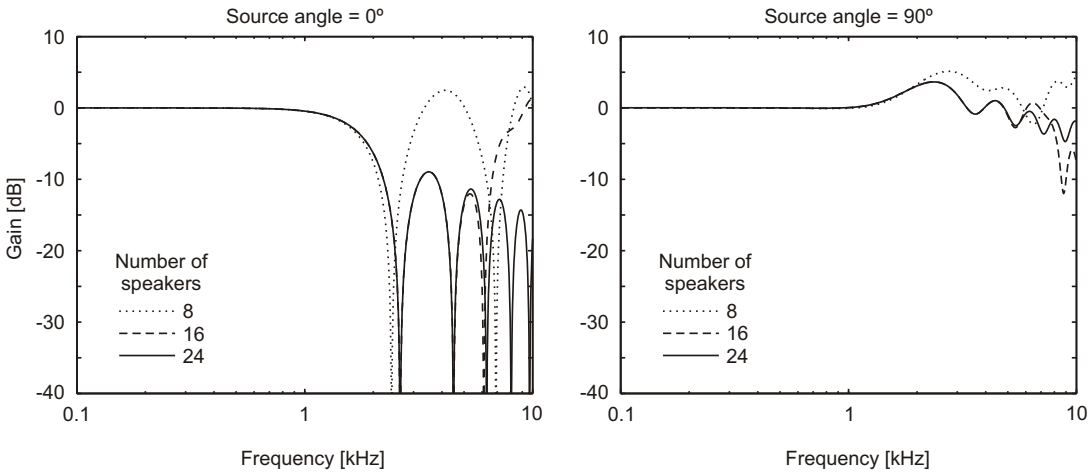

Fig. 7: Frequency responses for plane wave sources at 0° and 90° using 3rd order ambisonic encoding with different speaker numbers ($r = 0.1$ m).

rays, this results in the impulse responses illustrated in figures 5 and 6 being sampled at discrete points corresponding to the relative propagation delay from each speaker to the listening position. Keeping the number of speakers to the minimum, $(2M + 1)$, ensures that even in the 0° case where the impulse response of figure 6 is at its broadest, the discretized version will have most of the energy concentrated in just one or two impulses (depending on the orientation of the speaker array). This will lead to spectral impairment in the form of interferometric notches, but will not lead to the low-pass filtering that would be caused by the impulse responses illustrated in figure 6. Filtering effects will only be perceived if energy is distributed over a greater number of impulses as happens for higher speaker densities.

Figure 7 illustrates this phenomenon using 3rd order example arrays with different speaker densities. It can be seen that higher speaker densities cause more pronounced filtering effects at 0° source angle. However, at 90°, the effect is lessened significantly. The reason is that although energy is being distributed among several impulses, these impulses are clustered much more closely in time. This phenomenon can be predicted using figure 6. In terms of the frequency response, a closer time-distribution leads to a higher cut-off frequency, i.e. reduced spectral impairment.

5. SPEAKER DISTRIBUTION

5.1. Non-uniform Distribution

If the requirement that the speaker array is uniformly distributed is relaxed, it is possible to selectively increase the speaker density in the lateral regions without adversely affecting the system's frequency response. The examples presented in figure 7 show that much greater speaker densities can be tolerated for sources at 90° than at 0°. It follows that the speaker density can be increased around the ±90° areas as long as it is kept to a minimum around 0° and 180°.

If the degree of non-uniformity is increased too far, unwanted audible artifacts will occur because the trajectory of the energy vector magnitude will become increasingly distorted. This leads to a perceptually unbalanced soundfield.

In practice, it has been found that a good compromise is to use around twice the minimum number of speakers, i.e. $\sim 2(2M + 1)$. The extra speakers should be concentrated in the region between ±30° from the extreme left and right positions (±90°). If the speakers can be freely positioned (e.g. if using model-based HRTF data), a cubic distribution

function has been found to work well to smoothly distribute the speakers. If using a sampled HRTF set, only discrete angles will be available. For example, using the CIPIC HRTF database [12], a suitable speaker geometry for a 3rd order decoder is illustrated in figure 8. Ideally, the speaker density would have been greater around ±90° but sufficient HRTFs are not available using this data-set unless interpolation techniques are employed.

5.2. Ambisonic Decoder Design

Providing the non-uniformity of the speaker distribution is kept to a moderate level, as recommended in the previous section, an effective ambisonic decoder can be designed using the pseudo-inverse technique [13]. Non-uniform distributions generally lead to a non-uniform energy vector magnitude when plotted relative to source angle. In this case, the effect is an emphasis of the energy vector around the lateral regions.

In general use, the difference in energy vector magnitude between front/back and left/right locations is not readily perceived. However, when a sound source is slowly panned around a complete circle the irregularities in the energy vector trajectory can cause subtly audible artifacts. These can be suppressed by means of decoder optimization techniques described in previous works [6][9].

Example decoder coefficients in MATLAB [14] format can be found on the author's web-page at: www.eee.bham.ac.uk/collinst/binaural.

5.3. Subjective Listening Tests

At the time of writing, informal listening tests have been conducted involving the subjective comparison of three binaurally rendered virtual speaker arrays for 3rd order ambisonic decoding:

- Uniformly distributed 8 speaker array
- Uniformly distributed 16 speaker array
- Non-uniformly distributed 16 speaker array (using speaker positions from figure 8).

The results of these experiments confirm that the 8 speaker array suffers from a lack of definition in the lateral regions. Sound sources placed in the region roughly between ±50° from the extreme left or right angles, were hard to localize and were perceived as being diffuse compared to the well focused sound from front and rear locations. The uniformly distributed 16 speaker array, on the other hand, gave good localization performance for all angles. There was, however, a distinct coloration to sound sources from the front or rear. The non-uniformly distributed array gave good localization performance at all angles, again, but with no perceivable coloration. The timbral quality of the sound was not completely consistent when a source was panned around a complete circle but it is expected that further optimization of the decoder design will address this relatively subtle remaining artifact.

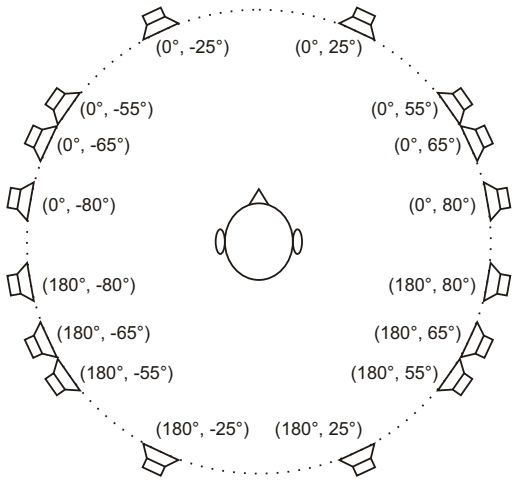

Fig. 8: 16 speaker array using speaker angles available in the CIPIC database. Speaker locations are shown in the format: $(elevation, azimuth)$ using the CIPIC coordinate system.

6. CONCLUSIONS AND FURTHER WORK

In much of the literature concerning ambisonic decoders, optimization is based on the magnitude and direction of the energy and velocity vectors. This is perfectly appropriate since the ILD and ITD do constitute the primary cues for binaural localization of sound sources. However, accurate localization also requires appropriate pinna filtering to be applied in order to disambiguate the source location from the

cone of confusion. This is especially so in the lateral regions where the ILD and ITD vary very little and the listener relies on pinna filtering information all the more.

In this paper, it has been demonstrated that extra virtual speakers can be added to a binaural decoder algorithm in the lateral region in order to enhance the accuracy of the pinna filtering but without the associated spectral impairment that would normally be associated with larger array sizes. Time-domain analysis of the impulse response of ambisonic systems was used to inform the design. The results of initial informal listening tests are very encouraging. The non-uniform distribution exhibits improved localization without perceptual spectral impairment. This improvement comes at no extra cost in the processing complexity of the final system.

Further subjective listening tests are planned for future work as well as extending the concept to three-dimensional ambisonic decoding. It is also possible that further work on the decoder optimization process may prove fruitful. An optimization process that includes spectral impairment and lateral localization performance as part of the overall fitness parameter (along with the velocity and energy vectors' magnitudes and directions) should be able to produce the optimal decoder coefficients for the complete 360° listening arena.

7. REFERENCES

[1] Michael A. Gerzon. Periphony: With-height sound reproduction. *J. Audio Eng. Soc*, 21(1):2–10, 1973.

[2] Audun Solvang. Spectral impairment of two-dimensional higher order ambisonics. *J. Audio Eng. Soc*, 56(4):267–279, 2008.

[3] Michael A. Gerzon. Criteria for evaluating surround-sound systems. *J. Audio Eng. Soc*, 25(6):400–408, 1977.

[4] Matthias Frank, Franz Zotter, and Alois Sontacchi. Localization experiments using different 2D ambisonics decoders. In *25th Tonmeistertagung VDT International Convention*, November 2008.

[5] Günther Theile and Georg Plenge. Localization of lateral phantom-sources. In *AES 53rd Convention*, Zurich, Switzerland, March 1976.

[6] Aaron J Heller, Eric M Benjamin, and Richard Lee. A toolkit for the design of ambisonic decoders. In *Linux Audio Conference*, 2012.

[7] C P Brown and R O Duda. A structural model for binaural sound synthesis. *IEEE Transactions On Speech And Audio Processing*, 6(5):476–488, 1998.

[8] Mark A. Poletti. A unified theory of horizontal holographic sound systems. *J. Audio Eng. Soc*, 48(12):1155–1182, 2000.

[9] Bruce Wiggins, Iain Paterson-Stephens, Val Lowndes, and Stuart Berry. The design and optimisation of surround sound decoders using heuristic methods. In *Conference of the UK Simulation Society*, pages 106–114, 2003.

[10] Adam McKeag and David S. McGrath. Sound field format to binaural decoder with head tracking. In *Audio Engineering Society Convention 6r*, 8 1996.

[11] Theodore Rivlin. *Chebyshev Polynomials*. Wiley, 1990.

[12] V.R. Algazi, R.O. Duda, D.M. Thompson, and C. Avendano. The CIPIC HRTF database. In *IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics*, pages 99–102, 2001.

[13] Aaron Heller, Richard Lee, and Eric Benjamin. Is my decoder ambisonic? In *Audio Engineering Society Convention 125*, 10 2008.

[14] MATLAB. *version 7.4 (R2007a)*. The MathWorks Inc., Natick, Massachusetts, 2007.