Audio Engineering Society

Convention Paper

Presented at the 129th Convention
2010 November 4–7 San Francisco, CA, USA

Why Ambisonics Does Work

Eric Benjamin[1], Richard Lee[2], and Aaron Heller[3]

[1] Surround Research, Pacifica CA, USA
ebenj@pacbell.net

[2] Pandit Littoral, Cookstown, AU

[3] SRI International, Menlo Park, CA, USA

ABSTRACT

Several techniques exist for surround sound, including Ambisonics, VBAP, WFS, and pair-wise panning. Each of the systems have strengths and weaknesses but Ambisonics has long been favored for its extensibility and for being a complete solution, including both recording and playback. But Ambisonics has not met with great critical or commercial success despite having been available in one form or another for many years. Some observers have gone so far as to suggest that Ambisonics can't work. The present work is intended to provide an analysis of the performance of Ambisonics according to various psychoacoustic mechanisms in spatial hearing such as localization.

1. INTRODUCTION

The authors of this paper have been involved with Ambisonics for many years and that involvement has been rooted in good experiences with the system. It is thus something of a surprise when others report that 'Ambisonics doesn't work'. The following criticisms of Ambisonics have been published [1,2].

(1) "The initial spike in the right ear is positive whereas the first spike in the left ear is negative. This is caused by the fact that loudspeakers that are opposite each other in the listening space in a 1st-order Ambisonics system are opposite in polarity."

(2) "The interaural time differences that occur with real sources are eliminated in the Ambisonics system. This is caused by the fact that the soundfield microphone cannot detect time of arrival differences because it is in one location."

(3) "a demonstration of four-loudspeaker Ambisonic recordings played in an anechoic chamber yielded an auditory impression that was almost totally within the head."

These criticisms will first be addressed briefly, one at a time, and then the underlying issues will be dealt with in further detail.

We would like to stress that the issue here is not to rebut the three criticisms listed above, but rather to use them as a stepping-off point to further address vitally important issues having to do with the way in which surround sound systems reproduce audio.

2. CRITICISMS ADDRESSED

1) The first criticism has to do with the polarity of the signals that arise at the listener's ears. "The initial spike in the right ear is positive whereas the first spike in the left ear is negative". This is a critical point; while it is not possible for the signals at the ears to be identical to what is experienced in natural hearing, it is vital that they be as similar as possible to natural hearing.

The signals referred to in quotation (1) were derived by using Head Related Impulse Responses (HRIRs) from the MIT Kemar data set [3] to construct a binaural model of what happens in a reproduction scenario where an octagonal array of loudspeakers was used along with the appropriate HRIRs to construct the binaural signals which would have been generated if Kemar had been placed in the center of the array. This was done specifically for the case where a signal was reproduced from a direction directly to the right of Kemar's head.

That result is shown in Figure 10.155 from reference [1]. Figure 10.155 was resampled to acquire the original data, and that data appears as follows:

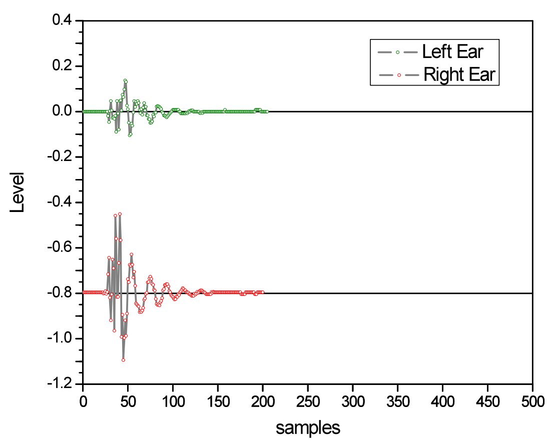

Figure 1: Ear signals in ambisonic reproduction with source at 90° right azimuth. Gray traces are from original figure, green and red dots are resampled points for left and right ears.

It's a little hard to see exactly what the polarity of each signal is in Figure 1. With the data from the resampling it is possible to make a more detailed analysis of what is happening in the ear signals from the octagon array. The impulse responses are quite complicated. In order to see what is happening at low frequencies, which is where the time behavior of the ear signals is most important, the impulse responses were low-pass filtered at 800 Hz using a 4^{th}-order Butterworth filter. They then appeared as shown in the following screen grab:

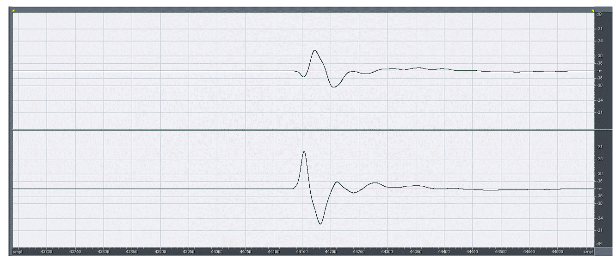

Figure 2: Impulse responses at left and right ears, with 8-loudspeaker reproduction system, low-pass filtered at 800 Hz.

With the high-frequency detail removed it is now clear that the left ear signal is similar to the right ear, but attenuated in level and delayed in time. Those signals can now be compared with the ear signals which are generated by a simple binaural reproduction scenario in which the HRIRs from the Kemar are reproduced without the interposition of the Ambisonic system. Those HRIRs look like this:

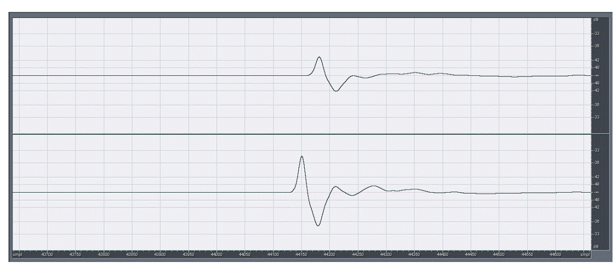

Figure 3: Impulse responses at left and right ears with binaural reproduction (natural hearing)

These two groups of HRIRs are actually quite close when examined with low-pass filtering. The ones from Ambisonic reproduction have more time delay between the left and right ears than do the ones from natural hearing.

2) The second criticism has to do with the presence of time differences in soundfield recordings. An Ambisonic soundfield recording, whether it is made with a Soundfield microphone or in a studio using panners, is one in which the soundfield at a point is represented by signals that represent the pressure and the three directions of the particle velocity at a single point in the reproduction venue. As such, it is clear that there are no explicit time differences in sounds coming from different directions. Time differences should arise at the listener's ears, and not in the recording itself.

The ITDs can be measured by filtering the impulse responses from Figure 1 with a 250 Hz one-third octave band-pass filter. Then the relative times of arrival at that frequency can be seen, as in the following figure. At this point the ear signals look almost identical. But if the time scale is made to be finer, then any differences in time of arrival will become visible:

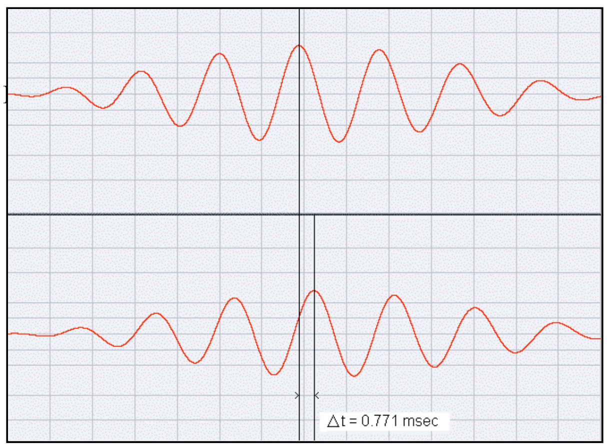

Figure 4: Ear signals after band-pass filtering, with fine time scale.

At this time scale it can be seen that the right ear signal does indeed arrive after the left ear, just as it should! The time delay is about 0.77 msec, which is about correct for a source directly to the left of the listener [6]. This will be investigated in more detail in section 3.

The third criticism was that:

3) "a demonstration of Ambisonics yielded an auditory impression that was almost totally within the head."

To address this criticism we must first understand what is meant by "within the head". Such perception of localization as being "in the head" is familiar to headphone listeners. It sometimes also occurs when listening to other forms of audio reproduction, but is almost never experienced in natural hearing.

Sound sources which are near the listener's ears produce ear signals (the individual acoustic signals at the entrance to the listener's ear canals) which are more strongly different than do distant sources. At low frequencies the levels between the two ears almost never differ by more than a few dB. If the source is close, such as when an insect perches on the listener's ear, then the difference in level can be quite large.

There are three suggested mechanisms which may cause in-head localization:

1. The sound images move with the listener's head, and are therefore interpreted as being a part of the listener's head.
2. The sound images aren't accompanied by the known reverberation characteristics of the environment, and are thus interpreted as not being a part of the outer environment.
3. Localization cues, such as ITD and ILD are not consistent with external position of the source, and therefore must be very near or inside the head.

It is well known that one of the principle mechanisms by which humans localize sound is that, at low frequencies, the sound from a single source arrives at one of the listener's ears earlier than at the other. That Interaural Time Difference (ITD) is the principle cue for identifying the direction of a low-frequency source. Above about 800 Hz, auditory localization switches over to using Interaural Level Differences (ILDs), which will be discussed later. What an audio system should do, however, is not necessarily record those ITDs but rather to record signals that produce the proper ITDs in the reproduction venue.

The important point here is to realize that the time differences occur when the listener's head interacts with the sound field, whether it is in the original performance venue or at the point of audio reproduction. An ideal audio system might be expected to reproduce the sound field exactly, in which case the listener's head would have the same interaction with the field as in the performance venue, and would hear exactly the same thing. We have only imperfect, non-ideal audio systems to use.

Therefore, an audio system should reproduce ITDs that are very similar to what would have been experienced in the performance venue. Reproduction of ITDs will be explored in the following section.

3. REPRODUCTION OF ITDS

One way to test a surround sound system is to investigate its performance in a virtual system in which binaural hearing is used to measure what happens, both for the recording space and for the reproduction space. This is exactly what was done in reference [1] and it is a very powerful technique. It is well known, for instance, what the changes are in ITDs for sources at various azimuth angles and ranges. If a (virtual) binaural recording were to be made of sound sources at different directions in the recording space then that recording could be used as a benchmark against which a binaural recording in the reproduction space could be compared. What is proposed, then, is that a virtual binaural recording and a virtual Soundfield recording will be made in the recording space and then a second binaural recording will be made in the reproduction space. Ideally, then, the ITDs embodied in the two binaural recordings would be similar or identical.

As a further simplification, the diffraction of a sphere will be used as a model of the Head Related Transfer Functions (HRTFs) since they embody most of the diffraction related changes that occur at the listener's ears. This has previously been shown to work well for modeling the basics of localization [3]. A spherical head model has the significant advantage that it's always the same, every time it is calculated.

The specific head model used here is for a sphere with a diameter of 20 cm, with the ears located at ±100° with respect to the front. For each azimuthal direction there is a unique response to each of the model's two ears, with attendant differences in times of arrival and spectral differences between the two ears.

Looking just at the time of arrival differences, the ITD was calculated at 5 degree azimuth intervals and on one-third octave intervals from 100 Hz up to 800 Hz. This was done by calculating the Impulse Responses from the source to each ear [3] for each direction, filtering them to the frequency band in question using one-third octave filters, and then measuring the time delay between the two signals that remain after filtering. The resolution in the time domain is somewhat limited by the sampling interval and the method used for measuring zero crossings. That introduced a quantization of the time interval of about 1/8 of a sample at a 48 kHz sample rate, or about 2.6 μsec. The ITD data calculated in this way are shown in the following figure:

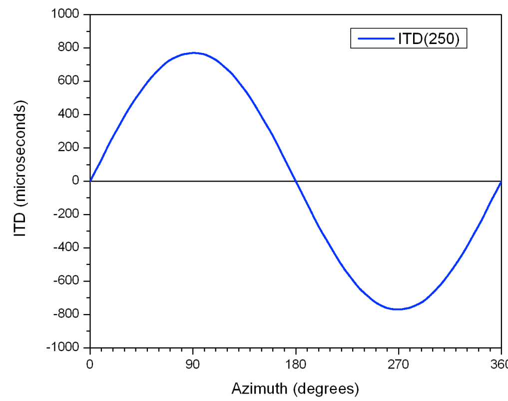

Figure 5: ITDs in natural hearing, spherical head model

The ITD varies almost sinusoidally, although the maximum ITD occurs at about 89 degrees, not 90 degrees, because the ears are located back of center. The data shown here are for 250 Hz, but the result was essentially the same at all frequencies.

If the azimuthal sound sources in the performance venue were to have been recorded by a Soundfield microphone, it would be according to the first-order Ambisonic encoding equations. For a sound source of magnitude S at azimuth θ and elevation φ:

$$W = S\frac{\sqrt{2}}{2}$$

$$X = S\cos\theta\cos\phi$$

$$Y = S\sin\theta\cos\phi$$

$$Z = S\sin\phi$$

The direction and intensity are embodied in four similar signals whose relative levels contain the information about the source strength and direction. If that signal is transmitted to the reproduction venue, then reproduction is accomplished by applying an Ambisonic decoder which is appropriate to the particular array of loudspeakers which exists there.

Such decoders have been discussed widely elsewhere. The important points are that the decoder is designed such that:

1) The particle velocity is reproduced with correct magnitude and direction

2) The energy vector is reproduced with correct direction and the maximum possible magnitude

The first requirement has been shown to be equivalent to reproducing the correct ITDs. The requirement for reproducing the correct ITDs can be tested by creating a virtual Ambisonic reproduction system and then examining the ear signals of a spherical head model placed at the center of the loudspeaker array. That is what was done by Martin [1], using a horizontal octagonal loudspeaker array, and will be repeated here, but with the difference that ITDs will be calculated from the ear signals.

The decoding equations [4] used in this experiment at low frequencies are as follows:

$$\begin{aligned}
S1 &= 0.1768W + 0.2310X + 0.0957Y + 0.0000Z \qquad (1)\\
S2 &= 0.1768W + 0.0957X + 0.2310Y + 0.0000Z\\
S3 &= 0.1768W - 0.0957X + 0.2310Y + 0.0000Z\\
S4 &= 0.1768W - 0.2310X + 0.0957Y + 0.0000Z\\
S5 &= 0.1768W - 0.2310X - 0.0957Y + 0.0000Z\\
S6 &= 0.1768W - 0.0957X - 0.2310Y + 0.0000Z\\
S7 &= 0.1768W + 0.0957X - 0.2310Y + 0.0000Z\\
S8 &= 0.1768W + 0.2310X - 0.0957Y + 0.0000Z
\end{aligned}$$

The 'Z' signal is not used since the octagonal loudspeaker array has no loudspeakers out of the horizontal plane.

Each loudspeaker can now be treated as a separate source, each with its own HRTFs to the head's ears. The total ear signal at each ear is the sum of the signals from the ambisonic decoder for each of the eight loudspeakers, weighted by the HRIR for the direction of the loudspeaker.

In the following figure, the ITDs calculated from the Ambisonic reproduction scenario are compared to those from the natural hearing case (Figure 1).

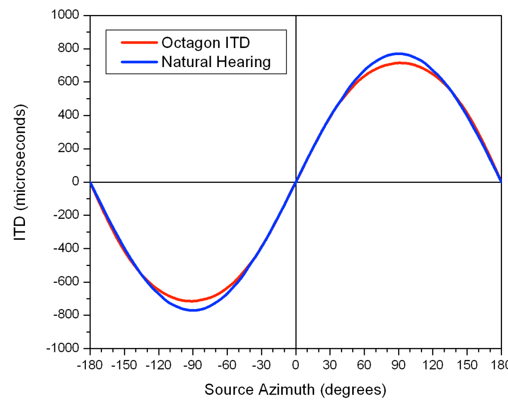

Figure 6: ITDs in Ambisonic reproduction compared to ITDs in natural hearing, spherical head model

It can be seen that the ITDs from natural hearing are reproduced quite closely, although not exactly. The closeness of the two results is quite satisfying.

It has been shown that Ambisonic decoders can reproduce the correct velocity vectors for a wide variety of loudspeaker array shapes. It would be expected from that fact that they should reproduce the correct ITDs, too, independently of the shape. The ITDs for a spherical head model were calculated for a diamond, square, and octagonal array, and are shown in the following figure:

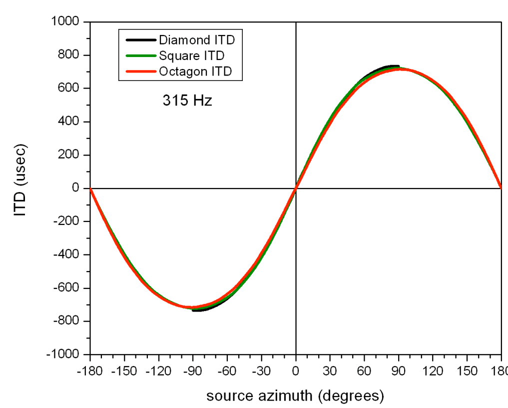

Figure 7: ITDs produced for a diamond, square, and octagonal array.

The ITDs are nearly identical regardless of the loudspeaker array shape chosen.

3.1. Analysis of Pair-wise Panning ITDs

The same analysis as is presented above for Ambisonic reproduction can also be done for pair-wise panning. Pair-wise panning is the same method as is normally done for frontal stereo using a panpot. In pair-wise panning (PWP) the image is moved from left to right across the frontal stage by varying the proportion of the two loudspeakers which encompass the intended direction. Only those two loudspeakers are active for sound sources in the front quadrant. Likewise, the image is moved along the sides by varying the proportion of the signal fed to the loudspeakers which encompass the side directions. The way in which the ratio is varied is by taking the sine and cosine of the intended angle, but this doesn't matter greatly since it is the ratio that is important and the images are usually placed by ear. This can be thought of as varying the amount of a particular HRTF that is associated with a particular loudspeaker location.

The individual ear signals can be calculated by summing the proportionate amount of the two active loudspeakers filtered by the HRTFs associate with the loudspeaker directions. As before, the ITDs can be calculated by filtering the resultant responses and comparing the time-of-arrivals at the two ears. When this is done for the case of pair-wise panning and a square loudspeaker array, the following ITDs result:

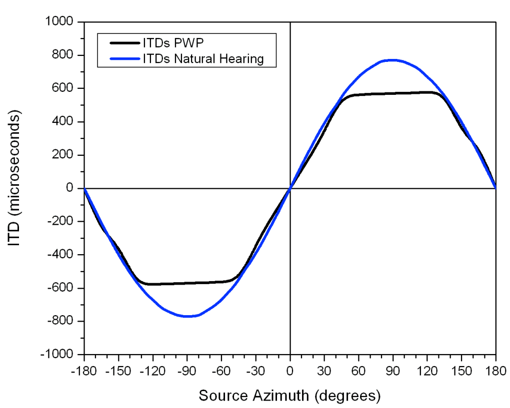

Figure 8: ITDs for pair-wise panning on a square loudspeaker array

The ITDs for pair-wise panning track ITDs for natural hearing fairly well for frontal sources, but fail completely for sources to the sides.

4. REPRODUCTION OF ILDS

At frequencies above about 800 Hz, depending on the individual, the perception of ITDs abruptly disappears. Instead, localization is governed by Interaural Level Differences (ILDs). The ILDs amount to spectral shaping caused by diffraction around the head, with high frequency components of sounds emphasized at the near ear and deemphasized at the far ear. These spectral changes are very complicated, even using a simple spherical head model.

The ILDs were calculated from the impulse responses from the source direction to the two ears, on one-third octave intervals and at 5 degree increments of azimuth.

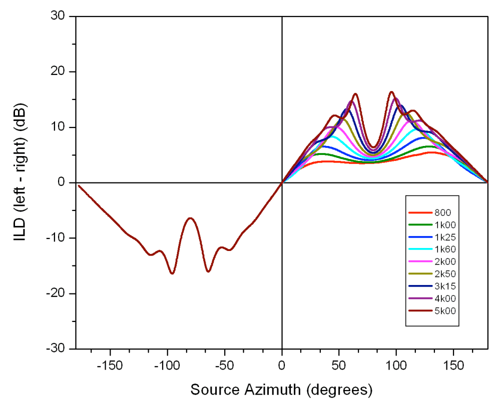

Figure 9: ILDs in natural hearing, spherical head model

The large dip at azimuth of about 80 degrees is due to the so-called "bright spot" effect. The name refers to the increase in level for sound diffracted around the interfering object, in this case the head, at an azimuth such that the path length is the same in all directions around the head. Although this is primarily a feature of theoretical constructs it does also appear in measurements of ILDs for real heads.

Finally, following a similar procedure as before, the ILDs were calculated for the scenario where the spherical head model is placed within an octagonal loudspeaker array, with the loudspeaker signals derived using the Ambisonic decoding equations (eqns 1).

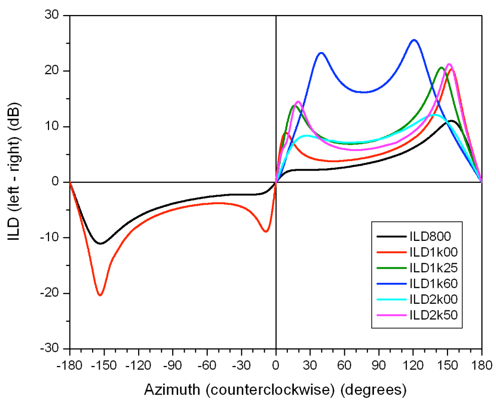

Figure 10: ILDs in octagonal array reproduction derived from spherical head model.

Although the general shape is similar, the ILDs calculated here seem to be generally larger than those for natural hearing, especially for azimuths near 0 and 180 degrees, and especially at frequencies near 1600 Hz. This results in a rapid change of ILDs for sounds at azimuths near 0. The very large ILDs at small azimuths are larger than what can be experienced in natural hearing (see Figure 3), except if the source were very close to the listener's head.

This is the reason for the in-head localization referred to in reference [2]!

Why are the ILDs for the octagonal array as different as they are from the ones for the natural hearing case? Referring back to section 3, the definition of an Ambisonic decoder is one that maximizes the Energy Vector $\mathbf{r}_E$. It is not expected that the reproduction will be exact and that the energy vector will be correct; the energy vector can't have its natural value of 1 unless the sound is coming directly from one of the loudspeakers.

The effect of the source azimuth on the ILD can be seen more easily if the ILDs are averaged over the frequency range of interest, in this case from 1 kHz to 3.15 kHz. It is easier then to see the general shape that the ITD curves take on as a function of azimuth.

Average ILDs for natural hearing are shown in the following figure:

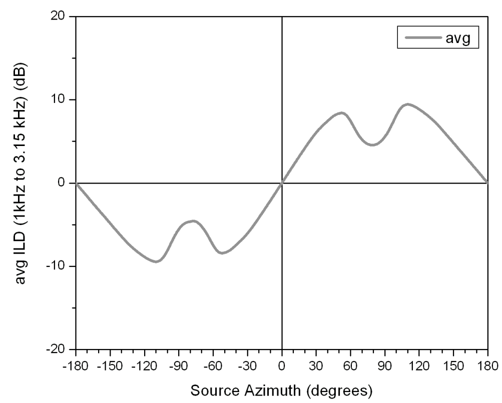

Figure 11: average ILDs for natural hearing, spherical head model.

At this point, in order to show that the results from the spherical head model are similar to what is experienced in real-world hearing, the same calculation was done except using HRIRs from the Listen HRTF database. That result is shown in the following figure:

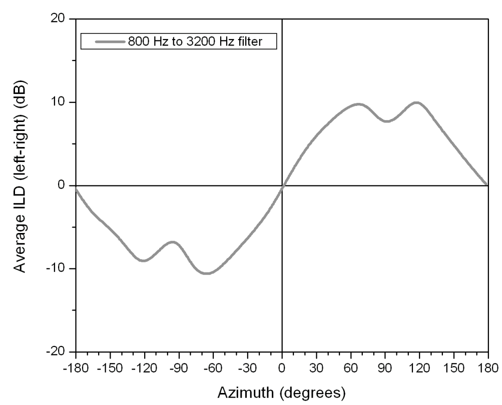

Figure 12. Average ILDs for natural hearing scenario, using HRIRs from Listen database.

The results for the HRIRs taken from the Listen HRTF database are extremely similar to those from the spherical head model, showing that the spherical head model does do a good job of representing the behavior of real heads, at least in the frequency ranges considered here.

The ILDs for Ambisonic reproduction were calculated using a similar method to what was described in section 2 for ITDs. The ILDs for an octagonal array are shown in following figure:

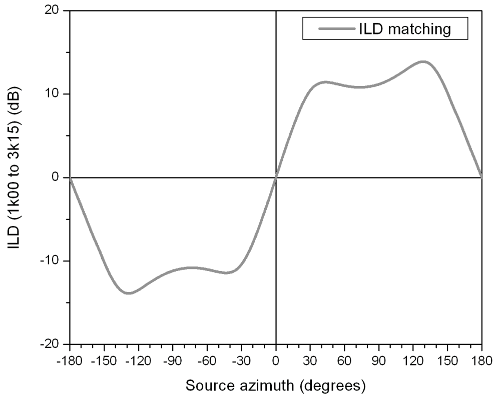

Figure 13: Average ILDs for octagonal array.

The ILDs for the octagonal loudspeaker system are larger than for the natural hearing case, that the 'bright spot' effect is smaller, and that the ILDs change more rapidly than natural hearing for azimuths near zero and 180 degrees.

Ambisonic reproduction is defined by Gerzon as having the decoder optimized to maximize the Energy vector $\mathbf{r}_E$ [8] at frequencies above 800 Hz. The following calculation was done to test the affect on the ILDs of maximizing $|\mathbf{r}_E|$.

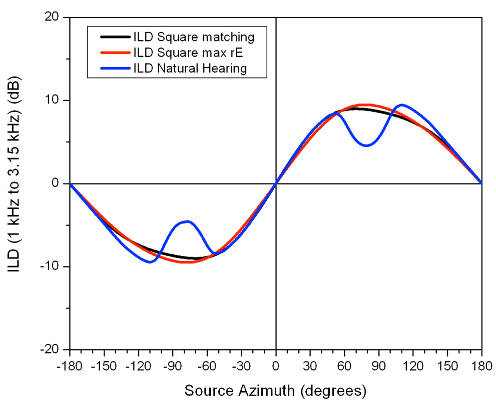

Figure 14: Average ILDs calculated for Ambisonics on a square loudspeaker array.

Max rE decoding does not appear to improve ILDs.

4.1. Analysis of Pair-wise Panning ILDs

As was done in section 3.1 for ITDs, the analysis of ILDs for pair-wise panning on a square loudspeaker array was performed for comparison with Ambisonic reproduction. The signals from the two loudspeakers which were active for any particular direction of source azimuth to the two ears of the spherical head model were summed to give the ear signals. The ear signals were filtered to a bandwidth of 1 kHz to 3.15 kHz to include the frequencies in which ILD localization is active, but exclude the frequencies where the pinna are significant in controlling localization.

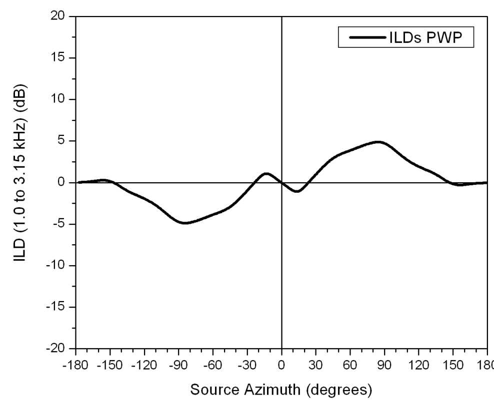

Figure 15: Average ILDs calculated for pair-wise panning on a square, using a spherical head model.

There are several significant discrepancies in the ILDs for pair-wise panning as compared to natural hearing. The first increment of panning away from 0° azimuth results in the ITD localization cues moving in the opposite direction as intended. At the same time, the ITDs, as shown in Figure 8 are producing cues that go in the correct direction. The result is that the localization is unstable for sound sources straight ahead, which has been noted in listening tests for pair-wise panned stereo sources when the loudspeakers are spaced by 90°, as they are in this case. The maximum value of ILD, for sources to the sides, is significantly less than in natural hearing, as seen in Figure 11.

5. REAL-WORLD EXPERIMENT

In natural hearing, an acoustic source emits sound that arrives at the listener's two ears at separate times and with differing levels depending on the direction of incidence. In Ambisonic sound reproduction a soundfield recording or a simulation is used to derive loudspeaker signals that are intended to reproduce a useful rendition of the original ear signals. In this latter case, a number of loudspeaker signals combine at the listener's ears to achieve the intended result.

How well this works in practice can be tested by experiment. The following experiment was performed. In an original 'performance' venue, a listener was exposed to several single acoustic events via signals reproduced from a loudspeaker, this representing the natural hearing case. The ear signals arriving at the entrance to the listener's ears were recorded using miniature microphones placed at the entrance to his ear canals. At the same time, the original sound field was recording using a Soundfield microphone. In the reproduction venue an array of loudspeakers was driven by loudspeaker feeds which were derived from the Soundfield recording using an Ambisonic decoder [7]. With the listener present, the same apparatus was used to record the ear signals and the sound field during the audio reproduction. Those recordings were then compared with the original recordings from the natural hearing case in order to evaluate the known psychoacoustic attributes.

The experiment is illustrated schematically in the following figure:

With the apparatus described above, two recordings are obtained in the performance venue, one a soundfield recording which represents the sound field at the listener's position, and the other a binaural recording which samples the signals at the entrances of the listener's two ears. The objective of the reproduction system is to reproduce those ear signals as accurately as possible, in particular the localization cues which enable the listener to locate the direction of the source. A perfect reproduction system would give precisely the same ear signals as were experienced at the performance venue and the same perceptual results as well.

The original acoustic source was a loudspeaker, which allowed the same acoustic event, whether from a recording or a synthetic test signal, to be produced repeatedly and identically. The soundfield microphone used was a well-calibrated Soundfield SPS422. The binaural microphones were Surround Research Bin2 microphones. The recordings were acquired using an IBM T42 laptop equipped with a Digigram VXpocket 440 4-channel sound card. The recorded files were then transferred to the reproduction system with four JBL LSR 6325 loudspeakers arranged in a rectangle of 4 meters diameter. The same Soundfield microphone and binaural microphones were used to monitor the reproduction system.

The Ambisonic recording was made after the binaural recording with the soundfield microphone located at the same position as the listener's head to ensure that the soundfield microphone received the same acoustic signals as the human listener. Each time the test signal was played was identical to the others because the acoustic source was a loudspeaker.

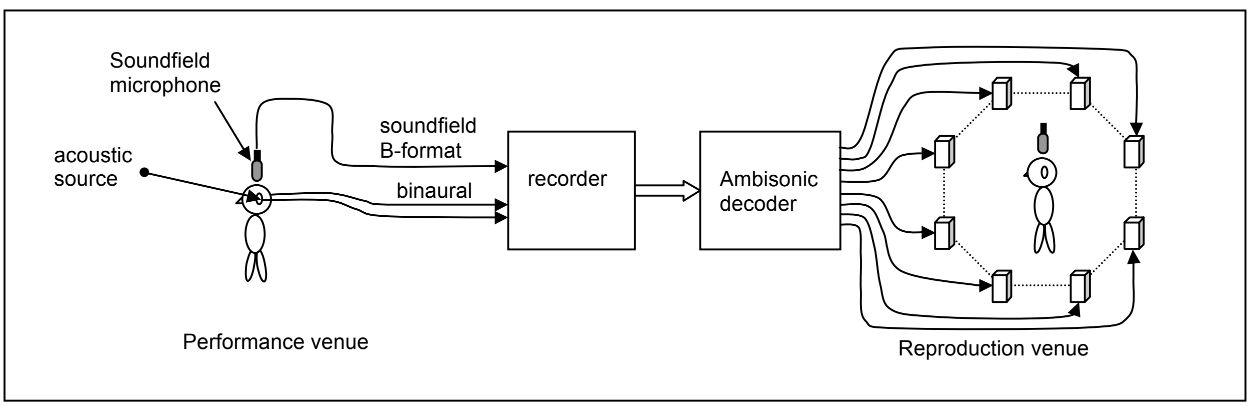

Figure 16: Ambisonic recording and reproduction with an octagonal loudspeaker array

5.1. Results from real-world experiment

It should be obvious to the reader that performing measurements on a real system comprised of real microphones, a real listener, and a loudspeaker system with real loudspeakers to determine the ITDs and ILDs is much more difficult to do accurately than it is to do the calculations for the ITDs for the spherical head model as is shown in Figure 6. There are additional sources of error having to do with the acoustic noise in the listening room and in the microphones, the subject's head will not be facing exactly forward, the subject's head isn't exactly symmetrical, the loudspeakers are not identical to each other, and their placement isn't exact.

The binaural recordings made in the listening room were processed using the same techniques used to calculate the ITDs and ILDs from the spherical head model.

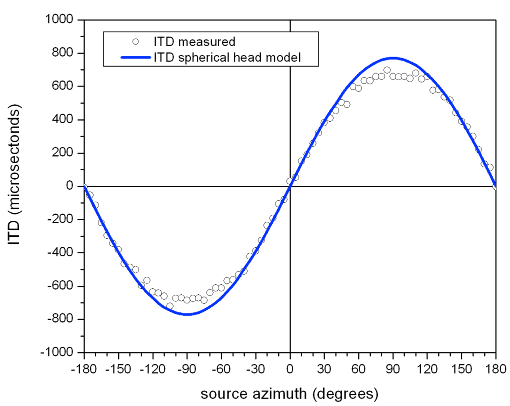

Figure 17: Measured ITDs in listening room with rectangular speaker array compared with ITDs calculated from spherical head model.

The ITDs measured from the binaural recording show significant amounts of error, and when compared with the ITDs from the spherical head model they are smaller. The average head diameter of the test subject was smaller than the 20 cm used in the spherical head model.

An attempt was made to calculate the ILDs from the binaural recording but errors in the recording prevented obtaining useful results.

6. ADVANCED AUDITORY MODELS

Another way assess the rendering quality of spatial reproduction is to feed synthesized ear signals into a sophisticated auditory localization model. Mariette has recently disclosed the result from an experiment where he uses Gaik's model to assess the rending quality of 2nd order Ambisonic reproduction [9]. Gaik's model [10] utilizes both ITDs and ILDs to produce a binaural activity map that shows the likely lateral perception of the sound.

The following is the binaural activity map that results from sampling a second order Ambisonic encode-decode process using a Dirac impulse signal at 36 discrete azimuths (10 degree steps) around the horizontal circle. The Ambisonic encode-decode used 128-point FIR shelf filters (with first-order and second-order pivot frequencies at 700Hz and 1200Hz respectively), rendered via 6 virtual speakers to binaural outputs utilizing measured HRTFs of a human subject.

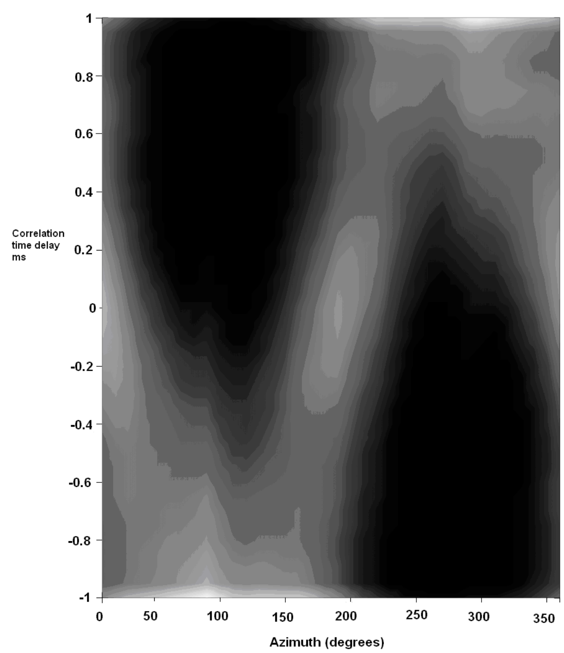

Figure 18: The binaural activity map produced by Gaik's algorithm from a test of 2nd-order Ambisonic reproduction. (courtesy Nick Mariette).

Similar features are visible here as in the other calculations; timing information that mimics natural hearing, as well as the "bright spot" phenomenon.

7. CONCLUSIONS

The introduction of this paper started out with several criticisms of Ambisonics. They were, stated briefly, that Ambisonics:

(1) produces reverse polarity signals at the listener's far ear for sounds to the side of the listener

(2) eliminates interaural time differences that occur with real sources

(3) produces "in-head" sensations

Through a series of investigations, using virtual recording and reproduction experiments, it was shown that the reverse polarity loudspeaker signals actually result in the correct polarity of signals at the listener's ears when examined at low frequencies. The polarity matters because the ear is sensitive to phase at low frequencies. Furthermore, when the Interaural Time Differences (ITDs) are calculated, it was shown that Ambisonics reproduces the ITDs almost exactly (Figure 6), and it does so regardless of the shape of the loudspeaker array (Figure 7) used for reproduction. Finally, when the Interaural Level Differences (ILDs) are calculated, they are similar in Ambisonics to what is experienced in natural hearing.

The "in-head" sensations are apparently produced by the cancellation of ear signals in the far ear under certain circumstances. This is a problem which occurs in any system that uses more than one loudspeaker to reproduce a single sound, and is associated with the magnitude of the Gerzon energy vector $\mathbf{r}_E$ being less than 1. If it were possible to increase the value of $\mathbf{r}_E$ then this problem will be diminished, and it is possible to do so by going to higher orders in the recording and reproduction system.

The results from the ILD calculations will be used in future work to attempt to further improve on present decoder implementations.

In summary, it was shown that Ambisonic reproduction, far from being subject to the criticisms listed above, is less subject to those problems than other forms of multichannel audio, notably pair-wise panning.

8. ACKNOWLEDGEMENTS

This work was inspired by the thoughtful publications of Geoff Martin and Floyd Toole, and fueled by the insights of Michael Gerzon, as expressed in the "General Metatheory of Auditory Localization". We thank Nick Mariette for allowing us to include his results.

9. REFERENCES

[1] Martin, G.; "Why Ambisonics Cannot Work", http://www.tonmeister.ca/main/textbook/node826.html

[2] Toole, F.; "Sound Reproduction; The Acoustics and Psychoacoustics of Loudspeakers and Rooms", Focal Press ISBN-10**:** 0240520092 (July 2008)

[3] http://sound.media.mit.edu/resources/KEMAR.html

[4] Gerzon, M., "General Metatheory of Auditory Localization", Preprint 3306, 92nd AES Convention, (Mar 1992)

[5] Furse, R.; "First and Second Order Ambisonic Decoding Equations", http://www.muse.demon.co.uk/ref/speakers.html

[6] Blauert, Jens, "Spatial hearing: the psychophysics of human sound localization", ISBN-10: 0262024136, MIT Press, (Oct 1996)

[7] Woodworth, R., Schlosberg, G., "*Experimental Psychology"*, pp. 349-361. Holt, Rinehard and Winston, NY (1962)

[8] Heller, A., Lee, R., Benjamin, E.; "Is my Decoder Ambisonic?", Preprint 7553, 125th AES Convention (Oct 2008)

[9] Mariette, N.; "Binaural activity map of Second Order Ambisonics with shelf filtering." Sonic Surrounds Blog http://blog.soundsorange.net/2010/09/

[10] Gaik, W; "Combined evaluation of interaural time and intensity differences: Psychoacoustic results and computer modeling," JASA, Vol. 94, No. 1 (July 1993)