Presented at the107th Conference of the Audio Engineering Society
New York, USA, 24 - 27 September 1999. Preprint no. 4996

Controlling Phantom Image Focus in a Multichannel Reproduction System

Geoff Martin[1], Wieslaw Woszczyk[2], Jason Corey[3], René Quesnel[4]

Multichannel Audio Research Laboratory, Faculty of Music, McGill University
Montréal, Québec, Canada

[1]martin@music.mcgill.ca, [2]wieslaw@music.mcgill.ca,
[3]corey@music.mcgill.ca, [4]quesnel@music.mcgill.ca

1 - Abstract

The focus of the phantom image of an anechoic sound source in a multichannel reproduction system can be determined through the manipulation of fundamental interchannel relationships. Through a series of listening tests, various functions controlling the channel amplitudes are evaluated for their accuracy and reliability in generating predictable impressions of image foci. The applicability of these findings to existing and future surround sound standards are discussed.

2 - Introduction

Over the past 100 years there have been numerous investigations into methods of reproduction and synthesis of an acoustic environment. More recently, such systems have attempted to create or re-create a virtual two- or three-dimensional soundfield generating direct, reflected and reverberant energies from the appropriate directions surrounding the listener using various methods. The methods by which phantom images are placed at a given location relative to the listener vary widely from simple pair-wise mixing through various algorithms which are used in an attempt to reconstruct a given soundfield [1] to processor-intensive convolution techniques using head-related transfer functions, head-tracking and crosstalk cancellation. As these systems evolve and different algorithms are employed to create synthetic environments we will require a greater understanding of the perception of phantom images in multichannel playback systems in order to generate appropriate image qualities for various sources.

This series of listening tests was performed to investigate the relationship between five different gain functions in an eight-channel loudspeaker configuration and the perceived sense of focus of

the resulting phantom images. The intention is to further our understanding of this relationship in order to increase our control of one of the many attributes of the perceived sound source. In the case of small sound sources, a system must be able to convey a very accurate and easily located sense of direction, whereas larger and more diffuse sources, be they direct or reflected in nature, should be more nebulous and difficult to locate.

In addition to investigating the association between the percepts of the phantom images and the various gain functions tested, interaural cross correlation (IACC) measurements were also conducted on the same set of gain functions. These measurements were performed to investigate the potential relationship between the subjects' perceptions of focus and the IACC's of the functions tested.

3 - System configuration

These findings were investigated using an eight-channel, radially symmetrical playback system shown in Figures 1 and 2. The configuration of this system is based on a hybrid of two ideals. The first is the requirement of a front centre channel, which has become an accepted practice in five-channel surround sound systems in order to stabilize centre images for multiple listener locations. The second is the use of real rather than phantom sources to ensure the stable and predicable location of lateral images. [2] When these are combined with the recommendation that loudspeaker apertures remain less than 60° in order to maintain stability of phantom image location, [2] the result is a configuration with loudspeakers placed at 0°, ±45°, ±90°, ±135° and 180° and equidistant to the listener.

The system was located in a room with a total volume of 82.57 m^3 and a modicum of acoustic treatment. Surface materials consisted of plaster on masonry walls, linoleum floor tiles on concrete and a suspended acoustic tile ceiling. In order to attenuate high frequencies in first-order reflections, particular ceiling tiles were replaced with mineral wool panels, high-frequency absorbers were hung on the walls and a rug was placed on the floor.

The sound source was an anechoic recording of continuous female speech routed through the eight analog outputs of a Mark of the Unicorn 2408 audio interface to a Yamaha 03D digital mixer. The outputs of the mixer were in turn connected to the eight Bang and Olufsen Beolab 4000 active two-way loudspeakers.

The entire system was calibrated in advance of the listening test to ensure that the time of arrival to the listening position was the same for each of the loudspeakers. In addition, the individual gains of the signal paths were matched using a static fader setting on the Yamaha 03D. This calibration was performed using a Brüel and Kjær 2012 single-channel audio analyzer with a 4191 free-field 1/2" microphone placed at the center of the circle delimited by the loudspeakers. Differences in both amplitude and propagation delays were measured using a series of Time Selective Response measurements on the 2012. The output buss faders of the console were used to align levels to match within a ±0.5 dB window. The distances between the loudspeakers and the listening position were also calibrated to propagation delay differences within ±10 μs during the same series of measurements.

A computer monitor with a maximum height of 103 cm (12 cm below tweeter level), keyboard and mouse were placed in front of the listening position for entry of responses during the listening tests. All calibration measurements were performed with the monitor in place in order to ensure that any detrimental acoustical effects of the screen's placement would not jeopardize the matching of the loudspeaker responses at the listening position.

4.1 - Test Procedures

The test was conducted using an automated software patch written by the principal author in Cycling '74's "MSP" DSP-development environment. All gain changes were performed by this system which was measured in advance of the listening tests to exhibit gain linearity within ±0.2 dB over the entire audio range. In addition, the software stored the responses of the test subjects for subsequent analysis.

The test group was comprised of three male students enrolled in the Music, Media and Technology doctoral program at McGill University. Although this is admittedly a relatively small number of subjects for a listening test, it was established that more participants would not be required due to the high degree of correlation between the independent responses of these first three subjects.

4.2 - Gain Functions

A total of five different gain functions based on three basic equations were tested in nine different locations. The functions tested were as follows :

4.2.1 - Pair-wise amplitude panning

This gain function is based on the cosine constant power amplitude panning algorithm used in panning controllers since the beginnings of stereo mixing consoles. In this case, the phantom or real image is generated by a minimum of one and a maximum of two loudspeakers, dependent on the intended direction to the perceived image. The equation used for this function is

$$G_n = \cos (2 (\theta - \phi_n)), [-\frac{\pi}{4} \le (\theta - \phi_n) \le \frac{\pi}{4}] \quad (1)$$

Where :

G_n is the gain of channel n

θ is the desired angle to the phantom image

ϕ_n is the angular location of loudspeaker n in the listening space

It should be noted that two modifications of the standard cosine gain function were employed due to the particular configuration of the loudspeaker placement. Firstly, the angle $(\theta - \phi_n)$ was multiplied by a factor of 2 due to the location of a loudspeaker at every 45°. Secondly, the equation was used for values of $-\frac{\pi}{4} \le (\theta - \phi_n) \le \frac{\pi}{4}$ only. For all other values, the gain G_n was set to 0.

As was the case with all functions tested, the particular function used for this investigation included a global gain correction applied to all channels, which was used to ensure a matched apparent amplitude of all stimuli.

4.2.2 - First-order Ambisonic

A first-order Ambisonic system employs both an encoding and decoding of the desired signal through the so-called "B-format" in which it is assumed that all sound sources occur on or within a unit sphere. It reproduces the zeroth- and first-order spherical harmonics of the original sound source, corresponding to the pressure and velocity of the encoded signal respectively.

A monophonic sound placed on the unit sphere will result in the following B-format channel signals : [3]

$$W = P_\psi \quad (2)$$

$X = P_\psi \cos \psi$

$Y = P_\psi \sin \psi$

Where :

W, X, and Y are the amplitudes of the three B-format channels

P_ψ is the pressure of the incident sound wave

ψ is the angle to the sound source (where 0° is directly forward) in the horizontal plane

It should be noted here that the fourth Z channel which has been omitted and the equations for deriving the X and Y channels have been simplified since we are assuming a two- rather than three-dimensional playback configuration.

These three channels are subsequently decoded according to the angular location of the loudspeakers as follows : [3]

$$P_n = \frac{W + 2X\cos\phi_n + 2Y\sin\phi_n}{N} \tag{3}$$

Where :

P_n is the amplitude of n^{th} loudspeaker

ϕ_n is the angle of the n^{th} loudspeaker

N is the number of loudspeakers

The decoding algorithm used here is one suggested by Vanderkooy and Lipshitz which differs from Gerzon's equations in that it uses a gain of 2 on the X and Y channels rather than the standard $\sqrt{2}$. This is due to the fact that this method omits the $\frac{1}{\sqrt{2}}$ gain from the W channel in the encoding process for simpler analysis. [3]

4.2.3 - Second-order Ambisonic

A second-order Ambisonic system builds on its first-order counterpart with the addition of second-order spherical harmonics, thus adding curvature information to the sound's wavefront. The W, X and Y channel remain the same, however two additional channels are encoded, represented by U and V:

$$U = P_{\psi} \cos 2\psi \qquad (4)$$
$$V = P_{\psi} \sin 2\psi$$

Where :

U and V are the amplitudes of the two additional B-format channels

P_{ψ} is the pressure of the incident sound wave

ψ is the angle to the sound source in the horizontal plane

These three channels are subsequently decoded according to the angular location of the loudspeakers as follows : [3]

$$P_n = \frac{W + 2X\cos\phi_n + 2Y\sin\phi_n + 2U\cos 2\phi_n + 2V\sin 2\phi_n}{N} \qquad (5)$$

Where :

P_n is the amplitude of n^{th} loudspeaker

ϕ_n is the angle of the n^{th} loudspeaker

N is the number of loudspeakers

4.2.4 - First-order gradient

Two other non-standard amplitude panning functions were tested for comparative purposes. The first was based on the first-order gradient function most commonly seen in descriptions of microphone polar patterns. The gains of the various channels followed the equation :

$$G_n = P + PG \cos(\theta - \phi_n) \qquad (6)$$

Where :

G_n is the gain of channel n

P is the pressure component where $0 \le P \le 1$

PG is the pressure gradient component and equal to $(1 - P)$

θ is the desired angle to the phantom image

ϕ_n is the angular location of loudspeaker n in the listening space

Note : In this test, both P and PG were 0.5

This function was tested to simulate the amplitude response which would result from a coincident microphone array comprised of matched cardioid transducers with included angles matching the angular locations of the loudspeakers — in this case, every 45°.

4.2.5 - Polarity-restricted cosine

The fifth and final function was developed to avoid a number of problems encountered in practice with the other algorithms tested. Although the pair-wise method described in section 4.2.1 provides a potential singularity in terms of the number of loudspeakers generating a given signal (and therefore would likely present the listener with the most focused signal), this can present a problem in a real listening space since the location of the actual loudspeaker can easily be determined by the listener. In addition, as has been noted in other studies, the negative-polarity gains produced by both first- and second-order Ambisonics systems can result in erroneous direction cues for listeners who are not positioned in the proper listening location — the so-called "sweet spot." [4]

The resulting algorithm was inspired by the concept of non-negative least squares, however, a much simpler algorithm was employed. A first-order gradient equation was used, restricting values to positive gains only — that is to say, any values produced by the function which were less than 0 were simply re-set to equal 0.

$$G_n = \cos(\theta - \phi_n), \left[-\frac{\pi}{2} \le (\theta - \phi_n) \le \frac{\pi}{2}\right] \quad (7)$$

Where :

G_n is the gain of channel n

θ is the desired angle to the phantom image

ϕ_n is the angular location of loudspeaker n in the listening space

4.3 - Presentation of Sound Stimuli

The functions were presented in bipolar pairs comparing two different gain functions with a matched phantom image location where the test subject was requested to indicate the most focused of the two sound stimuli. A copy of the control display seen by the test subjected can be found in Figure 3. All possible combinations of stimuli were tested, resulting in a total of 10 A-B comparisons presented in each of 9 intended image directions, from 0° to 180° inclusive, in steps of 22.5°. A numbered list of the bipolar pairs is presented in Table 1.

Pair	A	B
1	*a*	*b*
2	*a*	*c*
3	*a*	*d*
4	*a*	*e*
5	*b*	*c*
6	*b*	*d*
7	*b*	*e*
8	*c*	*d*
9	*c*	*e*
10	*d*	*e*

Where

a is the pair-wise amplitude panning algorithm

b is first-order Ambisonics

c is second-order Ambisonics

d is the first-order gradient function

e is the polarity-restricted cosine curve

Table 1 : List of bipolar pairs tested at each of 9 different angles.

The 90 pairs were presented in an order chosen randomly by the software at the time of testing. The response for a given pair was recorded when the subject moved to the subsequent exercise. The levels of all stimuli were matched and set to approximately
70 dB_{SPL} measured at the listening position.

5 - Results and Analysis

Before presenting the results of the listening test, some discussion is required regarding the definition of the term "focus" as it applies to images in a multichannel audio playback system. Following informal talks with the participants both before and after the test, it was established that a visual analogy, particularly focus as it applies to photography, cannot be used to adequately define the quality we are seeking. Specifically, the focus of an object in an optical medium is independent of size - it is a measurement of the perceived definition of the object's boundaries and details. Therefore, an object can be large yet focused. By comparison, the focus of the image of a single sound source in an audio playback system is related to its size. When a phantom image is larger or wider than the anticipated size of the actual sound source (in the case of this investigation, the human voice) the image is perceived as being unfocused — therefore the expected image size and perceived focus of the image are interdependent. In informal discussions following the test, all subjects agreed that any combination of the following three

attributes resulted in the perception of a more focused image : (1) narrower image width, (2) perceived increased distance to the image and (3) less spread of low frequency content.

The results of the listening test are presented in Figure 6. As can be seen in the plots for bipolar pairs 1 through 4, the pair-wise panning algorithm has a higher degree of perceived focus than all other algorithms. This was true for all angles with two notable exceptions being 22.5° and 157.5° for bipolar pair 2. In these two locations, subjects indicated that there was no difference in focus between the pair-wise panning and the second-order Ambisonic algorithm. This is an interesting anomaly deserving some discussion.

In the case of pair-wise panning, phantom image locations which match the angular loudspeaker position result in a single loudspeaker producing the program material. In a non-anechoic environment, this results in a very easily located image both in terms of angle and distance to the source, since there is no localization confusion caused by multiple coherent radiating sources as is the case with all other algorithms tested. When the phantom image is located between loudspeakers, there are a maximum of two radiating sources, the lowest maximum number for all five algorithms, again causing the least confusion. The second-order Ambisonic algorithm produces very focused images at all phantom image angles, however, when the image is located to the side of the listener, the sensation is one of a vacuum surrounding the ear to the opposite side from the image. This percept is similar to that caused by images produced by negatively correlated stereo loudspeakers with different levels. In addition, the perceived distance of the image appears closer to the listener than the images produced by all other algorithms. This perceived proximity is likely the result of a combination of an accurately synthesized wavefront and the use of anechoic source material, thus the least distance information is presented to the listener. Due to the perceived increased distance to the image generated by the pair-wise panning function, the focus appears to be greater than that of the second-order Ambisonics algorithm in most cases. The only exceptions occur in the instances where the pair-wise function is relying on two loudspeakers to produce the phantom image, and image is close to either 0° or 180°.

The results of bipolar pairs 5, 7, 8 and 10 indicate that the second-order Ambisonic and polarity-restricted cosine algorithms produce images with greater focus than either the first-order Ambisonic or the first-order gradient function. This trend was true for all angles tested.

Bipolar pair 9 shows that the images generated by second-order Ambisonic and the polarity-restricted cosine function have similar degrees of focus, however, this similarity is somewhat

misleading due to the fact that all phantom image angles are presented on the same chart. In fact, the second-order Ambisonic algorithm was found to have a greater degree of focus for images in the front and rear locations, whereas the polarity-restricted cosine function proved to generate more focused images for side positions.

Finally, the plot for bipolar pair 6 indicates that the focus of images generated by the first-order Ambisonic algorithm is marginally greater than those produced by the first-order gradient function. Again, this statement holds true for all angles tested.

Therefore, the functions listed in order of focus (from most to least focused) are as follows : (1) pair-wise panning, (2) polarity-restricted cosine and second-order Ambisonics, (3) first-order Ambisonics, and (4) first-order pressure gradient.

6 - Interaural Cross-correlation Measurements

The interaural cross-correlation of each of the 5 functions in all 9 image locations was measured using a Brüel and Kjær Model 2035 dual-channel signal analyzer with a 4100 Head and Torso Simulator. The simulator was positioned at the listening position, with the "ears" 1.15 cm above the floor, matching the tweeter height. Wideband white noise was used as the stimulus for this series of measurements, with the bandwidth limited between 22.4 Hz and 3.2 kHz at the input of the 2035. The measurements used a Hanning windowing function with 20 averages of the peak response.

The results of the IACC measurements are presented in Table 2 which lists the various values for all gain functions at all intended angles tested.

	0°	22.5°	45°	67.5°	90°	115.5°	135°	157.5°	180°
a	0.665	0.477	0.336	0.153	0.173	0.141	0.280	0.537	0.709
b	0.539	0.481	0.216	0.017	-0.093	-0.041	0.153	0.458	0.565
c	0.658	0.566	0.191	0.060	0.193	0.235	0.238	0.492	0.664

d	0.614	0.641	0.540	0.402	0.365	0.409	0.504	0.610	0.638
e	0.576	0.565	0.263	0.142	0.134	0.181	0.301	0.539	0.608

Table 2 : IACC for the five gain functions at all angles used in listening test. The list of gain functions *a* through *e* is presented with Table 1.

Some factors which can explain the results of the listening test are apparent. As expected, all five functions have similar IACC values for phantom images located directly forward and to the rear of the listener. Both Ambisonics algorithms are the only functions which approach complete inter-aural decorrelation for side images which may explain the "vacuum" effect noticed by the subjects. The first-order gradient function has the smallest range of values of all five functions tested — thus, when the intended direction to the phantom image is to the side of the listener, the IACC is greatest of the various functions, contributing to the spread of the image towards the front and back.

It is also interesting to note that the IACC measurements for the second-order Ambisonic algorithm is the only function whose values exhibit asymmetry across the frontal plane. Whereas all other functions have a minimum IACC level at 90°, the lowest value for this algorithm is achieved at 67.5° instead. This characteristic is perceivable to the listener, both as the increase in the aforementioned polarity-reversal effect, as well as a reduction in the apparent distance to the phantom image in this area.

Both the pair-wise panning and the polarity-restricted cosine functions exhibit similar IACC measurements for all angles in addition to being symmetrical across the frontal plane, thus exhibiting similar imaging characteristics as the image is panned around the listener.

7 - Conclusions

The results of this investigation suggest that the polarity-restricted cosine gain function should be used for the synthesis of phantom images in an eight-channel radially symmetrical loudspeaker configuration where a precisely focused image is desired. This method is preferred over the pair-

wise mixing in spite of the perceived greater focus of the latter due to the fact that informal investigations following this test show that the cosine function produces fewer changes in the perceived image qualities such as focus and distance to the phantom source with dynamic locations. This statement also holds true when comparing the polarity-restricted cosine function to the second-order Ambisonic algorithm.

In cases where a very unfocused image is desirable, the first-order pressure gradient function can be used, or modified as desired by changing the values of *P* and *PG*.

Further investigations are required to determine the robustness of the various algorithms in maintaining the desired source location for listener locations other than the designated "sweet spot." As was previously noted, it has already been established that some of the functions tested are inappropriate for inexact positioning of the listener. [4] Future studies will indicate the appropriateness of the various algorithms for more typical "real-world" usage.

Finally, it should be emphasized that the conclusions only hold true for an array of loudspeakers with matched levels and frequency responses. The use of unmatched transducers will invariably result in unpredictable imaging characteristics as the angular location of the phantom source is altered.

8 - References

[1] Nicol, R. and Emerit, M., (1999), "3D-Sound Reproduction over and Extensive Listening Area: A Hybrid Method Derived from Holophony and Ambisonic," *Proceedings of the Audio Engineering Society 16th International Conference*

[2] Theile, G. and Plenge, G. (1977), "Localization of Lateral Phantom Sources," *Journal of the Audio Engineering Society*, Vol. 25, No. 4

[3] Bamford, G. and Vanderkooy, J. (1995), "Ambisonic Sound For Us," *Audio Engineering Society Preprint 4138, 99th Convention*

[4] Dickins, G., Flax, M., McKeag, A., and McGrath, D. (1999), "Optimal 3D Speaker Panning," *Proceedings of the Audio Engineering Society 16th International Conference*

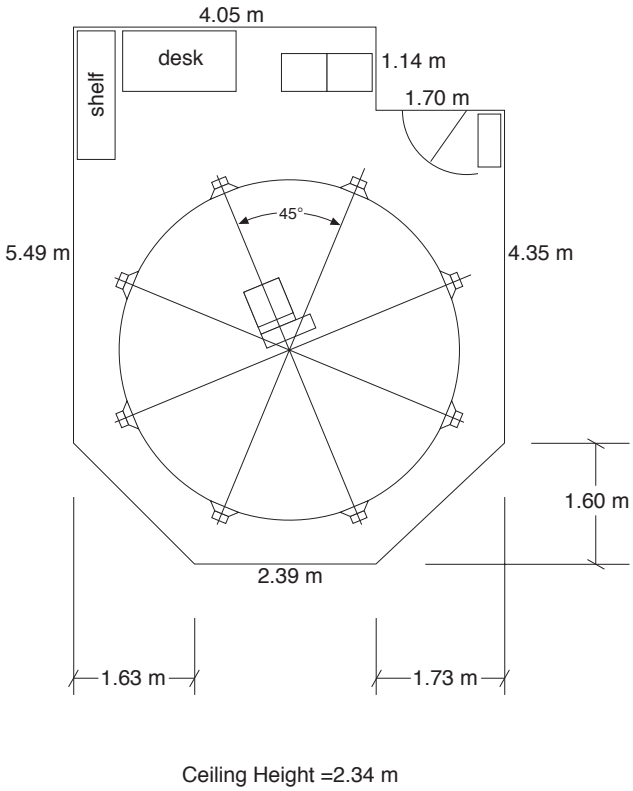

Figure 1 : Floor plan and loudspeaker configuration (to scale)

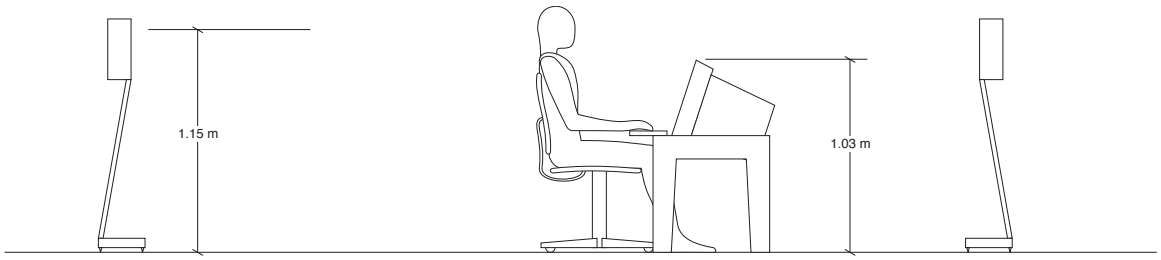

Figure 2 : Side view of loudspeaker configuration (side speakers omitted for clarity; to scale)

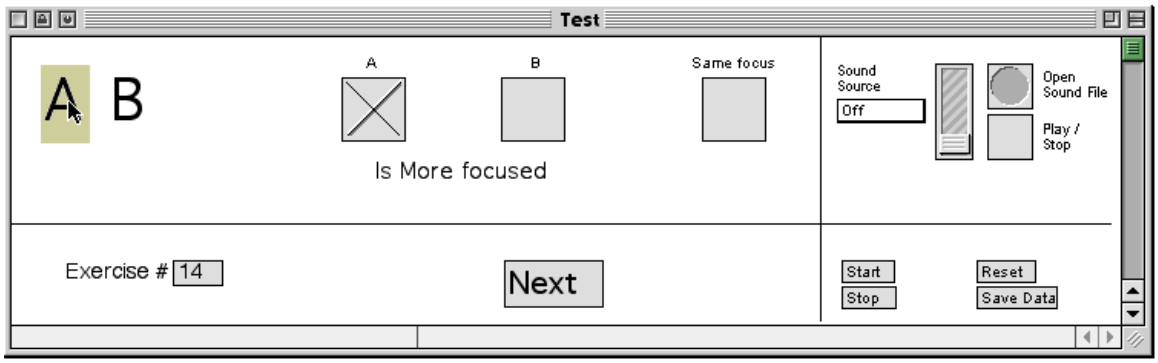

Figure 3 : Screen shot of display shown to test subjects.

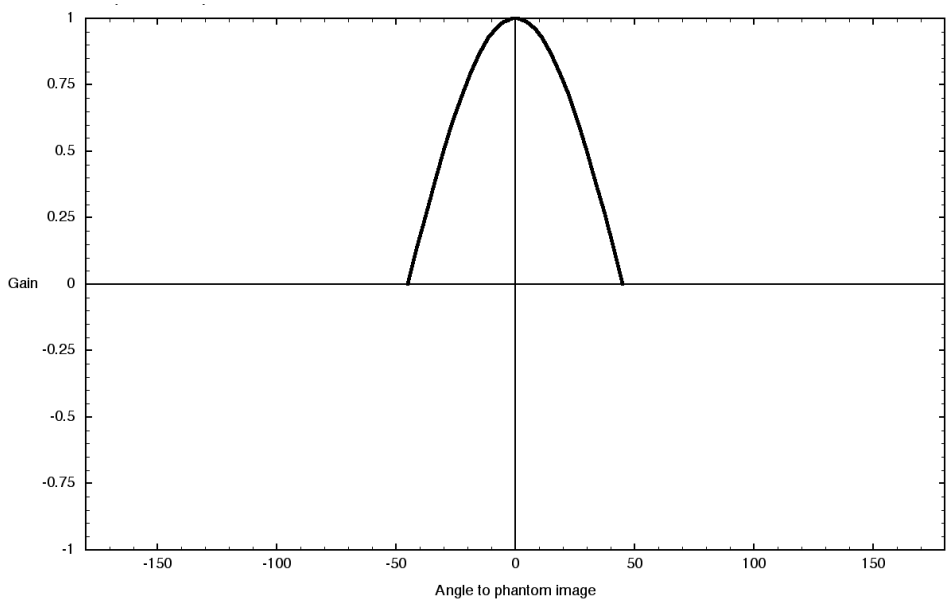

Figure 4a : Gain values for the centre loudspeaker vs. phantom image location for pair-wise panning algorithm. Values shown do not include gain correction applied to match summed power between the various algorithms.

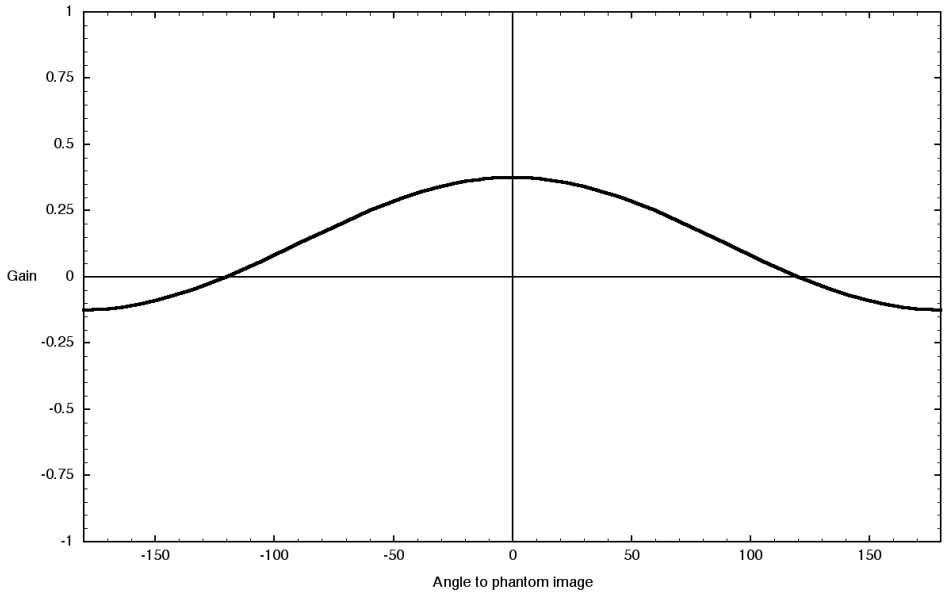

Figure 4b : Gain values for the centre loudspeaker vs. phantom image location for the first-order Ambisonic algorithm. Values shown do not include gain correction applied to match summed power between the various algorithms.

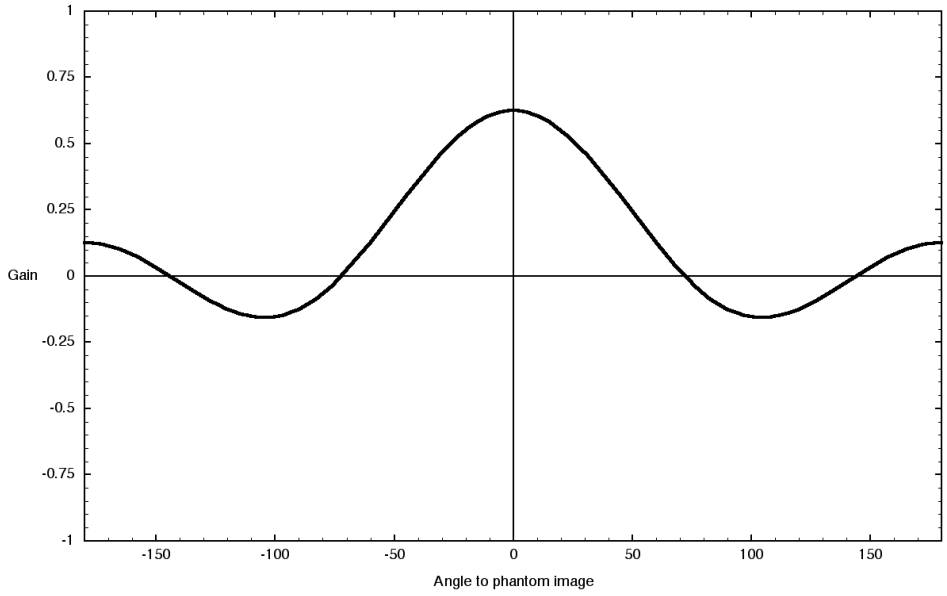

Figure 4c : Gain values for the centre loudspeaker vs. phantom image location for the second-order Ambisonic algorithm. Values shown do not include gain correction applied to match summed power between the various algorithms.

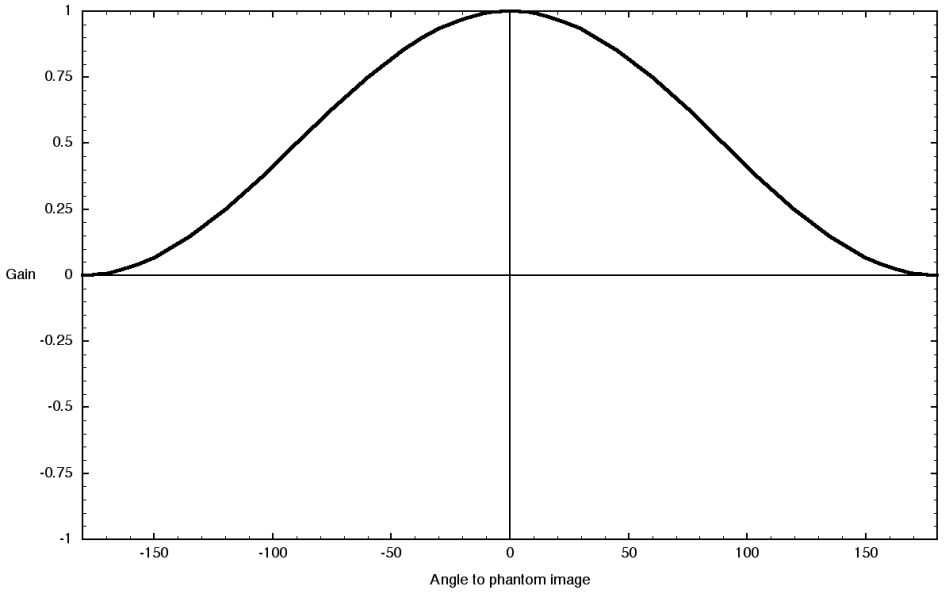

Figure 4d : Gain values for the centre loudspeaker vs. phantom image location for the first-order gradient algorithm. Values shown do not include gain correction applied to match summed power between the various algorithms.

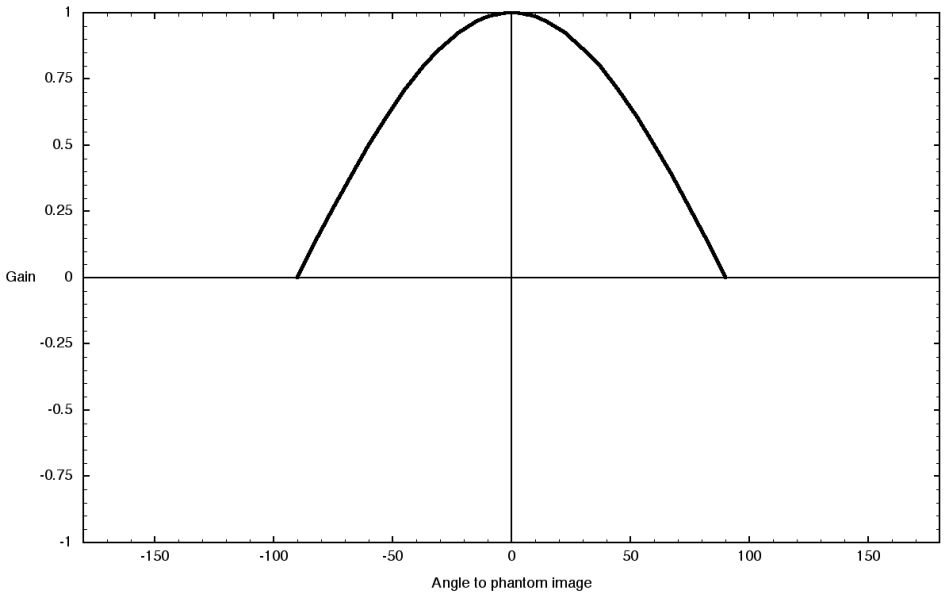

Figure 4e : Gain values for the centre loudspeaker vs. phantom image location for the polarity-restricted cosine algorithm. Values shown do not include gain correction applied to match summed power between the various algorithms.

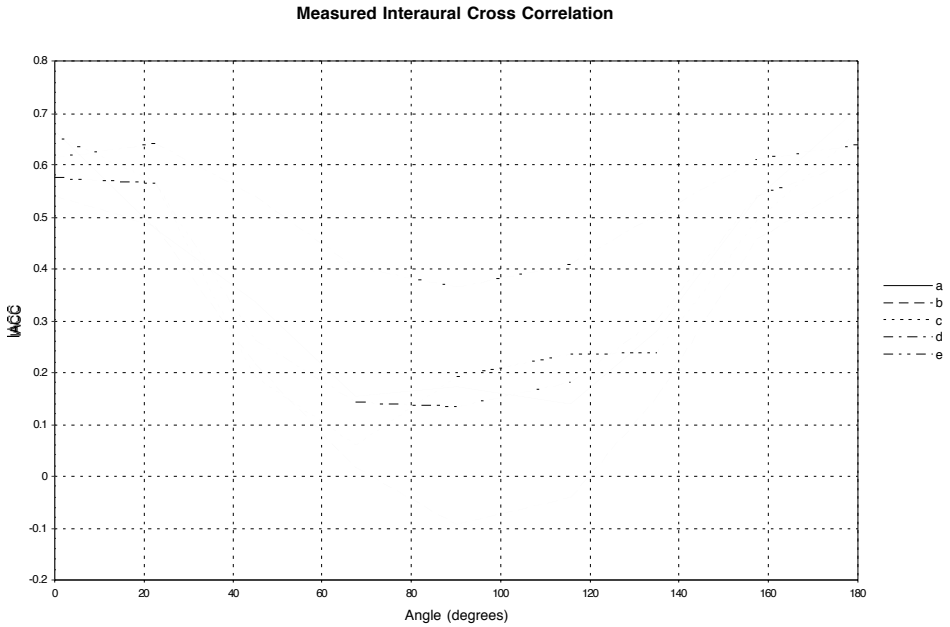

Figure 5 : Interaural Cross-correlation values for each gain function at all angles

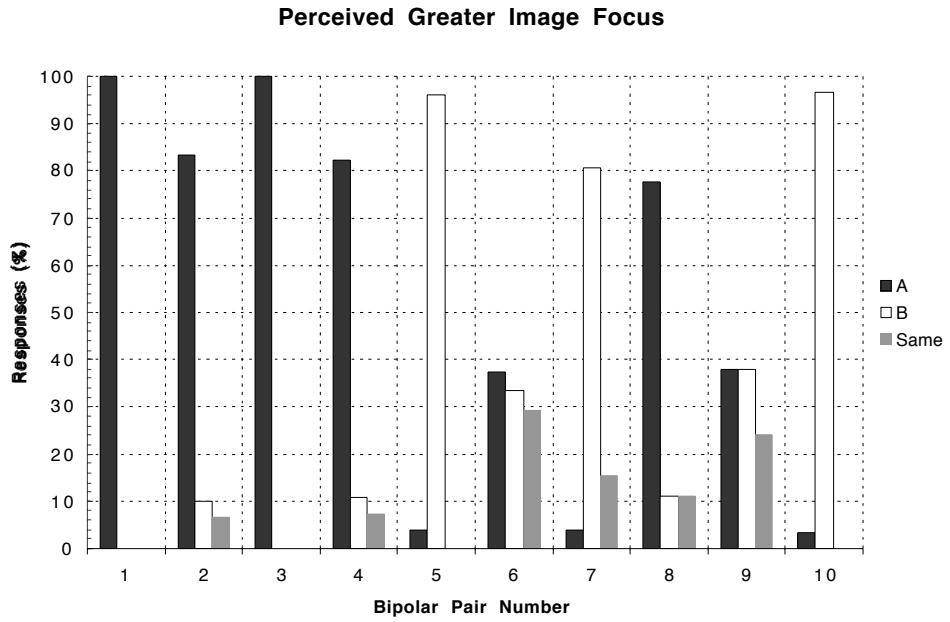

Figure 6 : Perceived Greater Image Focus for 10 bipolar pairs - all angles tested. Bipolar pairs are listed by number in Table 1.