Audio Engineering Society

Convention Paper

Presented at the 121st Convention
2006 October 5–8 San Francisco, CA, USA

Localization in Horizontal-Only Ambisonic Systems

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

[1]*Dolby Laboratories, San Francisco, CA 94044, USA*

[2]*Littoral Aficionado, Cooktown, Queensland 4895, AU*

[3]*Artificial Intelligence Center, SRI International, Menlo Park, CA 94025, USA*

Correspondence should be addressed to Eric Benjamin (emb@dolby.com)

Abstract

Ambisonic reproduction systems are unique in their ability to separately reproduce the pressure and velocity components of recorded audio signals. Gerzon proposed a theory of localization in which the human auditory system is presumed to localize using the direction of the velocity vector in the reproduced sound at low frequencies, and the energy vector at high frequencies. An Ambisonic decoder has the energy and velocity vectors coincident. These are the directions of the apparent source when the listener can turn to face it. Separately optimizing the low-frequency and mid/high-frequency operation of the reproduction system can optimize localization where the listener cannot turn to face the apparent source. We test the localization of horizontal-only Ambisonic reproduction systems using various test signals to separately evaluate low-frequency and mid-frequency localization.

1. INTRODUCTION

Two-channel stereo and, by extension, pair-wise panned multichannel audio, perform a primitive reconstruction of the original audio event. The relative intensities from two adjacent loudspeakers are varied in such a way that they sum to produce a pressure and a particle velocity vector pointing in the same direction as the original source. Because the pressures from the two loudspeakers add in a scalar fashion, while the velocities add vectorially, it is generally not possible to accurately recover the correct pressure and velocity at the listening position, except when the sound is panned to be at a particular speaker.

In contrast, Ambisonic audio reproduction systems use the full array of loudspeakers to control the sound field at the center of the array. It can be shown that it is possible, in principle, to reproduce the recorded sound field exactly at a single point in the center of the reproduction array.

While there is a great deal of material available in the open literature on the theory behind Ambisonics and there are commercial artifacts (Soundfield microphone, various decoders), very little has been published on the listening tests used to validate these designs. The principal contribution of this paper is to report on listening tests carried out where we compared a number of different speaker arrays and decoder designs. The main variables explored are the number and arrangement of the loudspeakers and the psychoacoustic models guiding the decoder design.

Gerzon has proposed a metatheory (a theory of theories) of auditory localization [1] in which he states that humans use many different mechanisms for auditory localization and that, except in cases where the cues are completely conflicting, the overall impression comes from majority decision.

He describes a hierarchy of models and for each, he derives a *localization vector* whose direction gives the predicted direction of the sound, and whose magnitude describes the stability of the localization. For a real, single-point source the magnitude of the localization vector is 1.0. If it is less or greater than 1.0 for a given decoder and speaker array, the perceived direction moves if the listener turns his head.

The two simplest, and possibly most important, models described are the *acoustic particle velocity model*, which corresponds to Makita's model [2], and the *acoustic energy-flow model*, which corresponds to De Boer's model [3]. Gerzon points out that practically all models of auditory localization, except the pinna coloration and impulsive (high-frequency) interaural time delay models, are special cases of these two models [1]. They are commonly referred to in the Ambisonics literature as the *velocity* and *energy* models, and the associated localization vectors the *velocity vector* and *energy vector*. We adopt this convention despite apparent contradiction that that energy is a scalar not vector quantity. They are broadly correlated with measurements of interaural phase difference (IPD) and interaural level difference (ILD), respectively. Blauert summarizes the results of a number of experiments in relating ITD and ILD to directional perception [4].

In applying these psychoacoustic models to the design of reproduction systems, Gerzon states [1]

> A decoder or reproduction system for 360° surround sound is defined to be Ambisonic if, for a central listening position, it is designed such that:
>
> i) velocity and energy vector directions are the same at least up to around 4 kHz, such that the reproduced azimuth $\theta_V = \theta_E$ is substantially unchanged with frequency,
>
> ii) at low frequencies, say below around 400 Hz, the magnitude of the velocity vector is near unity for all reproduced azimuths,
>
> iii) at mid/high frequencies, say between around 700 Hz and 4 kHz, the energy vector magnitude, r_E, is substantially maximised across as large a part of the 360° sound stage as possible.

Gerzon's metatheory of localization [5, 1] posits that the best possible localization for an array of loudspeakers occurs when the magnitude of the velocity vector is set to unity at low frequencies, and the magnitude of the energy vector is maximized at middle frequencies, with the transition between the two regimes taking place at a frequency between 300 Hz and 700 Hz [6]. The assumption is that, if the velocity localization vector and the energy localization vector are the same for a reproduced sound source as they are for a real sound source, then the perception is the same; the reproduced sound source sounds like the real one. Analysis shows that, although the velocity localization vector can be perfectly recreated by an appropriate array of loudspeakers surrounding the listener, the energy localization vector can be perfectly recreated only if the sound comes directly from a single loudspeaker. For sound sources in all other directions, the magnitude of the energy localization vector will be less than that of a real sound source. The Ambisonic system optimizes the energy localization vector in all directions, which necessarily compromises localization in the directions of the loudspeakers in favor of making the quality of the localization uniform.

The choice of the transition frequency between the two localization mechanisms is based on various published psychoacoustic experiments on localization [2, 4]. The experimental work performed for the present paper was designed to test the assumptions described above regarding optimizing the localization in the low- and high-frequency regimes, and the choice of the transition frequency between them.

2. DESCRIPTION OF EXPERIMENTS

2.1. Test Program Material

The first-order Ambisonic encoding equations are

$$\begin{bmatrix} \mathsf{W} \\ \mathsf{X} \\ \mathsf{Y} \\ \mathsf{Z} \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{2}}{2} \\ \cos\theta\cos\varepsilon \\ \sin\theta\cos\varepsilon \\ \sin\varepsilon \end{bmatrix} \mathsf{S} \qquad (1)$$

where θ is the azimuth relative to straight ahead, ε is the elevation, and S is the signal.[1]

Because the experiments reported on in this paper were restricted to horizontal-only reproduction systems, the Z signal is not used and $\cos\varepsilon$ is 1. Thus encoding of the test signals is simply a matter of scaling the test signal S down by 3.01 dB ($\frac{\sqrt{2}}{2}$) to create W, and scaling by the cosine (for X) and sine (for Y) of the direction θ from which the signal is intended to appear.

Program material originating acoustically is recorded with the Soundfield microphone[7] or an equivalent microphone array[8]. These microphone arrays have outputs corresponding to four coincident microphones: one omnidirectional microphone and three figure-of-eight microphones facing in the directions of the X, Y, and Z axes. Since the directivity of a figure-of-eight microphone is proportional to $\cos\theta$, the encoding described above is naturally achieved. Soundfield recordings made by one of the authors were used for the listening tests described below.

Additional test program material was realized by encoding test signals such as band-pass filtered noise and recordings of an alto female voice making vocal announcements. That encoding was done by taking a single mono recording and scaling it by the ratios described above before placing the signal into the tracks representing W, X, and Y.

2.2. Loudspeaker Arrays

Figure 1 shows two rectangular loudspeaker arrays with the same array radius of 2.00 meters, one square (2.83m $\times$ 2.83m) and the other elongated (3.72m $\times$ 2.15m) such that the ratio of length to width is $\sqrt{3}:1$. These two arrays are shown superimposed in such a way that it is possible to implement both at the same time for purposes of comparison. Both arrays have a radius of 2 meters,

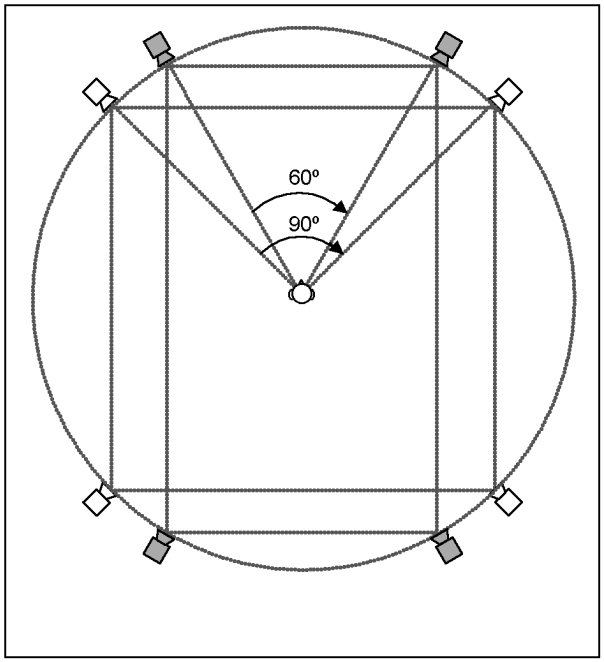


Figure 1: Square and rectangular ($\sqrt{3}:1$) Ambisonic arrays with 2 meter radius.

but the angle subtended by the front loudspeakers in the first array is 90° and in the second array it is 60°. The array with a ratio of $\sqrt{3}:1$ is of particular interest because both the front pair of loudspeakers and the rear pair of loudspeakers comprise a traditional stereo triangle. Since most domestic rooms are not square, it can be assumed that a rectangular array will fit into most rooms more conveniently than a square array. Classic Ambisonic decoders (e.g., [9]) are equipped with a *layout* control that allows the ratio of X and Y to be varied to accommodate rectangular arrays over the range of 2 : 1 to 1 : 2. The square array has uniform coverage in all directions, but theory shows that a rectangular array will have higher values of r_E in the direction toward the short side, which may possibly be an advantage for audio programs with a frontal emphasis.

To facilitate the direct comparison of the square and rectangular arrays, the decoding of the test programs was done beforehand and the decoded signals were compared

[1] The additional scaling factor of $\frac{\sqrt{2}}{2}$ in the W component is a historical artifact. It was added to improve the utilization of the dynamic range of recording media, based on the observation that the typical signal levels in the W channel are several dB higher than in X, Y, or Z.

using a multichannel file comparison utility. For both the square and rectangular arrays, the four decoded channels were inserted into an eight-channel file with the decoded signals for the square in the first four channels and the decoded signals for the rectangle in the second four channels. A schematic representation of the decoding and reproduction is shown in Figure 2. In this particular example, the comparison is between localization in square and rectangular arrays, but the same technique can also be used to compare between two different decodings for the same array. For example, two different sets of decoder parameters can be used, and the results can be compared in real time without the necessity for a real-time implementation of a decoder with continuously variable parameters. There are additional benefits to not doing real-time decoding. Every part of the process is under the explicit control of the experimenters and the resulting files can be tested, for instance to ensure that the spectrum of the speaker feeds has not been altered by the decoding process and that no clipping has occurred.

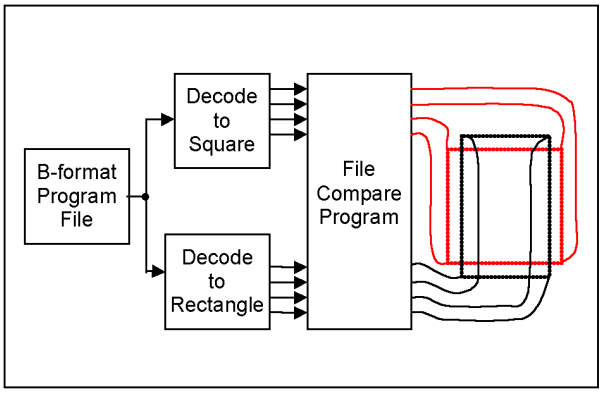


Figure 2: Arrangement for comparison of square decoding vs. rectangle decoding.

The principal limitations to using this technique are that there need to be enough channels of digital to analog conversion available, and enough loudspeakers to receive their signals. Comparisons of larger arrays, such as a hexagonal array vs. an octagonal array, will require very large numbers of speakers. Various additional loudspeaker array comparisons are shown in Appendix A.2.

2.3. Decoding Equations

In order to properly recover the horizontal Ambisonic components when the acoustic signals are summed at the center of the array, different decoding equations are required for the square and rectangular arrays.

The *Diametric Decoder Theorem* [10] states that the velocity- and energy-localization vectors coincide if

- All speakers are the same distance from the center of the layout.
- Speakers are placed in diametrically opposite pairs.
- The sum of the two signals fed to each diametric pair is the same for all diametric pairs.

When these conditions are met, we can design a decoder as follows.[2] Let n diametric speaker pairs lie in the directions

$$\pm(x_i, y_i, z_i) \tag{2}$$

for $i = 1, 2, \ldots, n$, then the respective speaker-feed signals are

$$\mathsf{S_i}^{\pm} = \mathsf{W} \pm (\alpha_i \mathsf{X} + \beta_i \mathsf{Y} + \gamma_i \mathsf{Z}) \tag{3}$$

where

$$\begin{bmatrix} \alpha_i \\ \beta_i \\ \gamma_i \end{bmatrix} = \frac{\sqrt{2}}{2} nk \left(\sum_{j=1}^{n} \begin{bmatrix} x_j^2 & x_j y_j & x_j z_j \\ x_j y_j & y_j^2 & y_j z_j \\ x_j z_j & y_j z_j & z_j^2 \end{bmatrix} \right)^{-1} \begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix} \tag{4}$$

with $k = 1$ at low frequencies, yielding the so-called velocity decode. Setting $k = \frac{\sqrt{2}}{2}$ and $\frac{1}{2}$, yields the energy and cardioid decodes, respectively. For horizontal layouts, all terms involving z are omitted (otherwise the matrix is singular) and $\gamma_i = 0$. Gnu Octave [13] code to numerically solve this is provided in the appendix.

In the rectangular case, if we let the angle subtended by the front speakers be 2ϕ, the analytic solution is

$$\alpha = \frac{1}{\sqrt{2}\cos\phi} \tag{5}$$

$$\beta = \frac{1}{\sqrt{2}\sin\phi} \tag{6}$$

where the growth in the Y-gain (β) relative to the X-gain (α) is needed to compensate for the growth of the rectangular speaker array in the x (front-back) dimension. As mentioned earlier, some hardware decoders provide an

[2] Strategies for designing decoders for speaker arrays that do not meet these conditions is covered in [11, 12]. Evaluation of such arrays will be a topic of a future paper.

approximation to this adjustment with a *Layout* control that ranges from an aspect ratio of 2 : 1 to 1 : 2, corresponding to a range of 2ϕ from 53° to 127°.

In practice, the speaker feeds are scaled to provide an exact reconstruction of the pressure at the center of the array, in this case by $\sqrt{2}/4 \approx 0.3536$. The coefficients used for the listening tests described in this paper are listed in Tables 1 and 2.

These basic decoding equations are the ones that satisfy the dual requirements of uniform coverage and recovering the correct magnitude of the sound pressure, p, and the correct magnitude and direction of the sound particle velocity, $\bar{v}$. Substituting the encoding equations into the decoding equations results in recovering the correct values for the pressure and the particle velocity at the center of the reproduction loudspeaker array.

However, reproducing the correct values for p and $\bar{v}$ does not necessarily give the best perceived localization, because the correct reconstruction is achieved over only a small area ($< \lambda/2$). If it were possible to recover the original wave field over a large area, nothing additional would need to be done. However, because the first-order Ambisonic system is unable to recover the original wave field over a large area, the Ambisonic technique is to exactly reproduce the velocity vector from the original location at low frequencies, and to maximize the energy vector at high frequencies.

The decoding equations used are

$$\mathsf{LF} = \sigma\mathsf{W} + \alpha\mathsf{X} + \beta\mathsf{Y} \quad (7)$$

$$\mathsf{RF} = \sigma\mathsf{W} + \alpha\mathsf{X} - \beta\mathsf{Y} \quad (8)$$

$$\mathsf{RR} = \sigma\mathsf{W} - \alpha\mathsf{X} - \beta\mathsf{Y} \quad (9)$$

$$\mathsf{LR} = \sigma\mathsf{W} - \alpha\mathsf{X} + \beta\mathsf{Y} \quad (10)$$

where σ, α, β are the values from Table 1.

The 'velocity' decoder equations for a hexagon are

$$\mathsf{S}_1 = 0.23570\mathsf{W} + 0.28868\mathsf{X} + 0.16667\mathsf{Y} \quad (11)$$

$$\mathsf{S}_2 = 0.23570\mathsf{W} + 0.28868\mathsf{X} - 0.16667\mathsf{Y} \quad (12)$$

$$\mathsf{S}_3 = 0.23570\mathsf{W} + 0.00000\mathsf{X} - 0.33333\mathsf{Y} \quad (13)$$

$$\mathsf{S}_4 = 0.23570\mathsf{W} - 0.28868\mathsf{X} - 0.16667\mathsf{Y} \quad (14)$$

$$\mathsf{S}_5 = 0.23570\mathsf{W} - 0.28868\mathsf{X} + 0.16667\mathsf{Y} \quad (15)$$

$$\mathsf{S}_6 = 0.23570\mathsf{W} - 0.00000\mathsf{X} + 0.33333\mathsf{Y} \quad (16)$$

where S_1 is the feed for the front-left speaker and then S_2–S_6 proceed clockwise around the array.

2.4. Distance Compensation

Distance compensation is a correction for a physical effect having to do with the radiation behavior of the loudspeakers. The near field has a "large velocity component out of phase with the pressure, [... this is] reactive energy which does not radiate outward." [14] This "large velocity component" is in fact in quadrature with the pressure, and as a result, the particle velocity at the listener's position is partly from the in-phase velocity radiation and partly from the reactive component. How near to the loudspeaker do these effects occur? The answer is wavelength dependent. The frequency at which the quadrature and in-phase components are equal is given by

$$f_c = \frac{c}{2\pi r} \quad (17)$$

where c is the speed of sound and r is the distance.

Thus the velocity component recreated at the center of the loudspeaker array will have a phase error that increases at low frequencies, and is greater for small systems with the listener close to the loudspeakers. This phase error must be corrected in order to ensure proper localization at low frequencies. This can easily accomplished by high-pass filtering the X and Y signals to bring them back into phase with the W signal. For a reproduction array with a 2-meter radius the high-pass filter should be at 27 Hz.

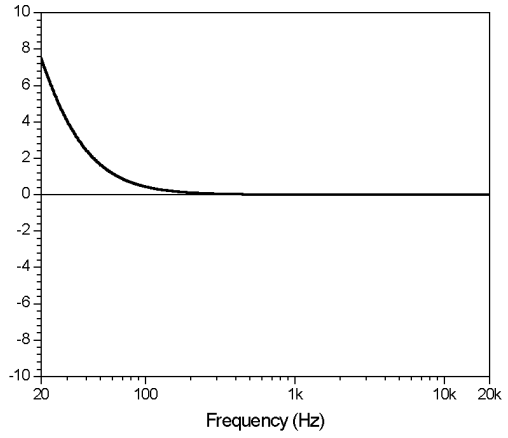


Figure 3: Phase (as time) between velocity and pressure components for loudspeakers at distance of 2 meters.

2.5. Velocity- and Energy-Localization Vectors

The velocity localization vector at the center of the re-

Aspect Ratio (X:Y)	Frontal Angle (deg)	X (m)	Y (m)	W-gain (σ)	X-gain (α)	Y-gain (β)
1 : 1	90.00	2.8284	2.8284	0.3536	0.3536	0.3536
$\sqrt{2}$: 1	70.53	3.2660	2.3094	0.3536	0.3062	0.4330
$\sqrt{3}$: 1	60.00	3.4641	2.0000	0.3536	0.2887	0.5000
2 : 1	53.13	3.5777	1.7889	0.3536	0.2795	0.5590

Table 1: "Velocity decoder" coefficients for rectangular arrays. These reproduce the exact pressure and velocity, but only over a very small area at high frequencies ($< \lambda/2$), and hence are suitable only for use at low-frequencies, say below 400 Hz.

production array is calculated by summing the velocity vector contributions from each of the loudspeakers. Because the Ambisonic decoding equations are derived to recover the velocity components exactly, the result is that the velocity vector always has unity magnitude and points in the direction of the intended source. Following Gerzon, the magnitude of the velocity vector, r_V, at the center of a speaker array with n speakers is

$$r_V \hat{r} = \mathrm{Re} \sum_{i=1}^{n} G_i \hat{u}_i / \sum_{i=1}^{n} G_i \qquad (18)$$

whereas the magnitude of the energy vector, r_E is computed by

$$r_E \hat{r} = \sum_{i=1}^{n} |G_i|^2 \hat{u}_i / \sum_{i=1}^{n} |G_i|^2 \qquad (19)$$

where the G_i are the (possibly complex) gains from the source to the i-th speaker, and $\hat{u}$ is a unit vector in the direction of the speaker. Computing the magnitude of the energy vector for all angles of azimuth, for a square loudspeaker array and two rectangular loudspeaker arrays, yields the graph in Figure 4. Examination of the polar plots of the energy vector magnitude versus direction for rectangles of various aspect ratios shows that rectangular layouts with various aspect ratios have higher values of r_E at the front and back, at the cost of having lower values of r_E at the sides. The value of r_E can also be examined as a function of the ratio of velocity to pressure, which is the fundamental decoder parameter.

In Figure 5 the value of r_E is plotted for rectangles of various aspect ratios and for a hexagon.

This confirms the observation from the polar plots that rectangles have greater values of r_E in the front than regular polygons, and suggests that if maximizing the energy vector magnitude is important for localization, the

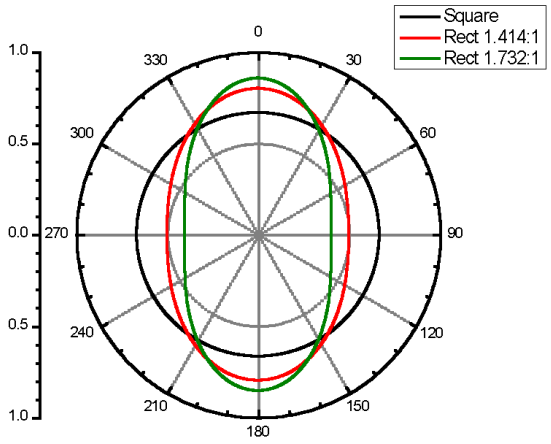


Figure 4: Magnitude of energy vector r_E as a function of the source angle for a square and two rectangles. The rectangular arrays have greater values of r_E at the short end, relative to square arrays.

rectangles would be better directly in the front than the square and hexagon. The square and hexagon have identical results, as will any regular polygon. It can also be seen that the ratio of velocity to pressure that maximizes r_E is different for rectangles with different aspect ratios.

The maximum r_E for rectangular loudspeaker arrays is explored further in Figure 6, which shows the optimum value of the velocity/pressure ratio for rectangles of different aspect ratios:

It can be seen that the value of the ratio of velocity to pressure that optimizes frontal r_E is 0.707 (the "energy decoder") for a square, and reaches a value of 0.89 for a rectangle with an aspect ratio of 2 : 1. Not only is the optimum value of the velocity to pressure ratio different for

Aspect Ratio (X:Y)	Frontal Angle (deg)	X (m)	Y (m)	W-gain (σ)	X-gain (α)	Y-gain (β)
1 : 1	90.00	2.8284	2.8284	0.4330	0.3062	0.3062
$\sqrt{2}$: 1	70.53	3.2660	2.3094	0.4330	0.2652	0.3750
$\sqrt{3}$: 1	60.00	3.4641	2.0000	0.4330	0.2500	0.4330
2 : 1	53.13	3.5777	1.7889	0.4330	0.2420	0.4841

Table 2: "Energy decoder" coefficients for rectangular arrays. These reproduce the pressure exactly, but the velocity is reduced by $\sqrt{2}$, which enlarges the listening area at mid-to-high frequencies. If no shelf filters are employed, this provides the best reproduction.

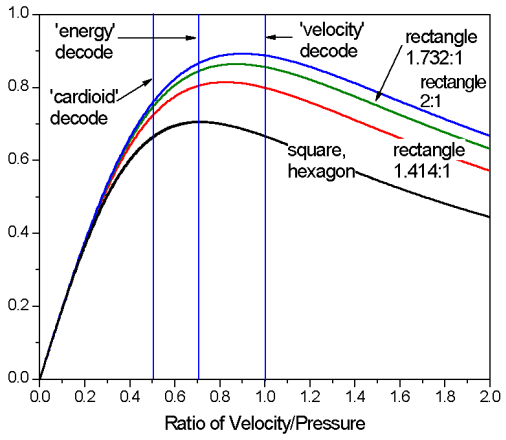


Figure 5: Maximum energy vector magnitude as a function of velocity/pressure ratio for various rectangles and hexagon.

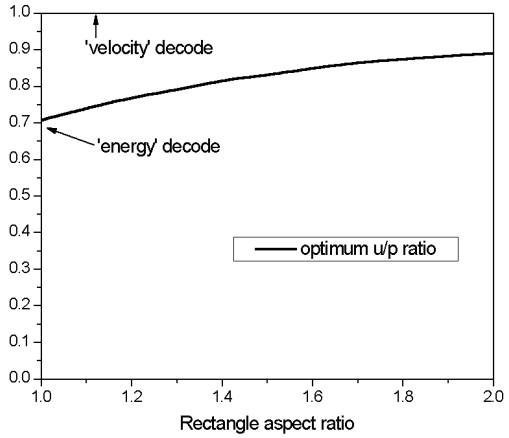


Figure 6: Velocity/pressure ratio for maximum energy vector magnitude at front.

rectangles with different aspect ratios, it is also different for different directions. That factor is not explored here.

What will be investigated is the audible localization quality of the frontal image for various loudspeaker arrays and decoder coefficients, specifically a square array, a $\sqrt{3}$: 1 rectangular array, and a hexagonal array.

Gerzon wrote that [5]

> "The ratio of the length of the above-defined energy vector to the total reproduced energy should ideally be unity; in practice the larger it is the better defined the sound image."

In summary, according to Gerzon's localization theory, a decoder achieves the best low-frequency localization by setting the magnitude of the magnitude of reproduced velocity vector to unity, while a decoder achieves the best middle-frequency localization by maximizing the magnitude of the reproduced energy vector. Optimizing both low-frequency localization and mid-frequency localization is achieved by the use of shelf filters.

2.6. Energy Decoding

Given that Gerzon's localization theory shows that the best localization at low frequencies is achieved by setting the magnitude of the reproduced velocity to unity, while the best localization at middle frequencies is achieved by maximizing the magnitude of the energy localization vector, it may be desired to design a decoder that max-

imizes that quantity without regard to the recovered velocity.

As is shown in Figure 5, such a decoder will be achieved by decreasing the ratio of velocity to pressure (that is, the ratio of X and Y to W) by 3.01 dB. In order to keep the perceived loudness constant when comparing *velocity* and *energy* energy decodes, we increased W by 1.76 dB (1.2247) and decreased both X and Y by 1.25 dB (0.8660). This gives different decoder coefficients relative to Table 1. The decoder coefficients for 'energy' decoding for various rectangular arrays are given in Table 2.

The 'energy' decoder equations for a hexagon are

$$S_1 = 0.2887W + 0.2500X + 0.1443Y \quad (20)$$
$$S_2 = 0.2887W + 0.2500X - 0.1443Y \quad (21)$$
$$S_3 = 0.2887W + 0.0000X - 0.2887Y \quad (22)$$
$$S_4 = 0.2887W - 0.2500X - 0.1443Y \quad (23)$$
$$S_5 = 0.2887W - 0.2500X + 0.1443Y \quad (24)$$
$$S_6 = 0.2887W - 0.0000X + 0.2887Y \quad (25)$$

These are the decoding equations that were used to perform energy decoding for the listening tests reported in Sections 3 and 4.

2.7. Shelf Filters

The change from 'velocity' Ambisonic decoding to decoding that is optimized at both low and high frequencies is accomplished by the use of shelf filters. Separate filters are applied to the W and X, Y signals (or, for periphonic reproduction, X, Y, and Z) to change the relative magnitude of their contributions, while keeping them in phase with each other. For the horizontal-only case, W is increased by 1.76 dB at high frequencies, while X and Y are decreased by 1.25 dB at high frequencies. The overall effect, then, is to have a 3.01 dB increase in the contribution of the pressure signal at high frequencies, while the spectrum of the energy is kept constant.

The shelf filters selected for the initial phase of this investigation are intended to mimic the performance of various commercially available Ambisonic decoders, which have a transition frequency of approximately 400 Hz. It is critically important that both filters are in phase with each other throughout the audio range. For the purposes of the listening tests reported in this paper, the shelf filters were implemented as finite-impulse-response filters of order 4096. The filter shape was within 0.01 dB of

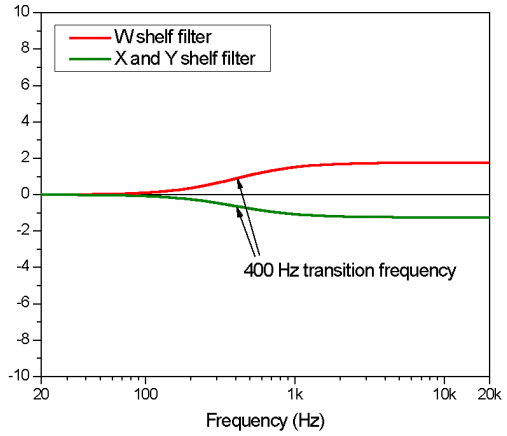


Figure 7: W and XY shelf filters for horizontal-only reproduction systems to be used only with the velocity decoding equations in Section 2.3.

a first-order IIR shelf filter with a 400 Hz transition frequency and the shelf gains specified above.

In this paper we follow the convention that speaker matrix gains are derived using the velocity matching criteria (i.e., $k = 1$ in Equation 4) and shelf filters are used to transition to energy-maximizing criteria above 400 Hz. An alternative approach is to use energy-maximizing criteria to derive the speaker matrix gains ($k = \frac{\sqrt{2}}{2}$) and use shelf filters to transition to velocity-matching criteria at lower frequencies. This approach has several practical advantages [15]. Furthermore, it has been suggested that it may be desirable to use yet another set of decoding criteria above 5 kHz. Therefore, when discussing a particular set of shelf filters, it is important to specify the speaker matrix with which they are to be used.

3. EXPERIMENTAL RESULTS

The initial listening tests were performed in an ordinary room, a domestic space in the residence of one of the authors. For reasons that are not known, those tests were not successful, in that good localization was not achieved. No experimental data are reported for those listening tests, but the authors intend to revisit whatever issues may have been responsible for that failure. The remainder of the listening tests took place in an acoustically treated professional listening room. The room measures 4.64 meters in width by 6.75 meters long, with the

ceiling at an approximate height of 2.64 meters. A plan view of this room is shown in Figure 8. The listener was

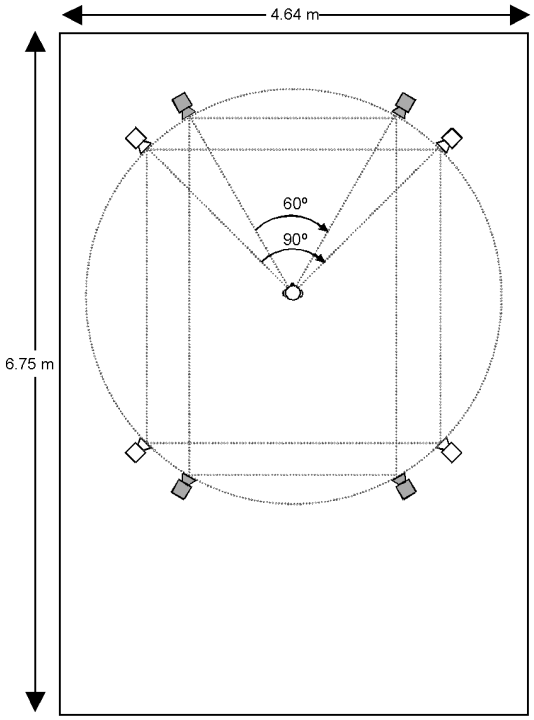


Figure 8: Listening room in which speaker array comparisons were made.

centered on the long axis of the room, which was necessary to keep the loudspeakers away from the side walls. The loudspeaker arrays were moved forward in the room, however, which kept the listener away from the geometrical center of the room. The loudspeaker locations were made to be within about 1 centimeter with respect to the desired theoretical locations. It is worth noting that small errors in the placement of the loudspeakers ($\approx 5-10$ cm) results in a shift in the tonal balance as well as a degradation of localization.

The loudspeakers used were JBL LSR25p powered monitors, mounted on stands with the acoustical center of the loudspeakers at 1.0 meter height, which is ear height for a listener seated in the chair used in this test. Eight nominally identical loudspeakers were drawn from a group of thirteen speakers utilized in a previous listening test. The frequency responses for the group of loudspeakers are shown in Figure 9.

A typical room response for these loudspeakers in the

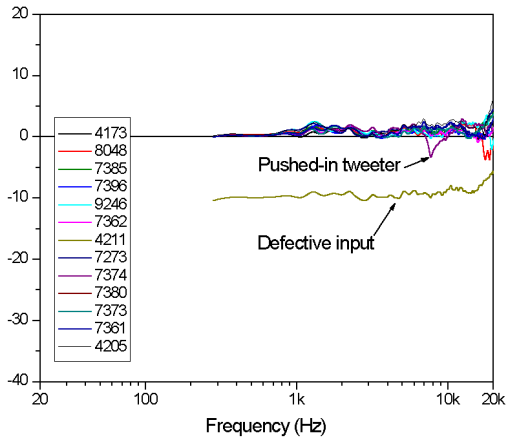


Figure 9: Frequency response of loudspeakers used for localization listening tests.

listening room is shown in Figure 10. This shows very

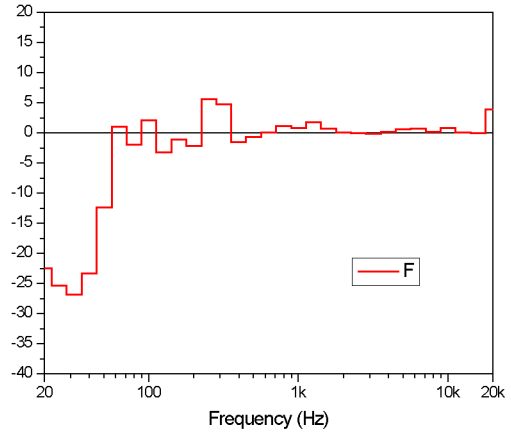


Figure 10: Room response of a single JBL LSR25p.

smooth response above 300 Hz. The room response below 300 Hz varied depending on the loudspeaker position.

The first listening tests compared the localization performance of a square array to that of a rectangular array with an aspect ratio of $\sqrt{3}:1$ (two 60° stereo setups back to back), as shown in Figure 8.

Three decoder formulations were also generated for each of the two layouts. These were a 'velocity' decoder in

which the original pressure and particle velocity are recovered exactly, an 'energy' decoder in which the energy vector magnitude is maximized by offsetting the relative gains of pressure and velocity by 3 dB, and a 'shelf' decoder in which the decoding conformed to the velocity coefficients at low frequencies and the energy coefficients at high frequencies, using the shelf filter described in Section 2.7 creating the transition between the two regimes at 400 Hz.

The various decoder configurations were tested using a file containing a single-sample impulse in W, then X, then Y. Each impulse is separated in time one second. This allows verification of gains and filter responses.

No attempt was made to make the tests blind or double blind. Listeners were aware of the decoder type and speaker array in use and free to listen as long as desired and switch between them at any time. Listeners were also free to move their heads and torsos, as well as move their chairs if desired. The floor was marked with the exact center of the array, so that the listener could return to the correct position.

The following recordings were used:

- 700 Hz broad band noise panned continuously around the array
- Voice announcements panned to the eight cardinal directions.
- Three-piece folk music recording, with the musicians arrayed around the microphone in all directions and relatively close to the microphone
- Classical string quartet recording in a fairly reverberant environment, recorded from approximately 2 meters away
- Classical chamber orchestra recording made in a 1200-seat hall with very good acoustics ($RT_{60} =$ 2.2s at midband), recorded 4-meters back and 2-meters above the conductor's head.
- Two minutes of applause from the above recording
- Outdoor recording of fireworks, both close and distant

The test subjects were experienced, critical listeners, accustomed to the sound of live music as well as high-quality audio reproduction systems. The subjects were asked to listen for the following:

- Directional accuracy of localization in each direction
- Perspective of localization in each direction (in head, near head, at speakers, beyond speakers)
- Compactness of the virtual images in each direction
- Static or dynamic "speaker detent" effects
- Overall tonal balance
- Changes in tonal balance with direction
- Reproduction artifacts, such as comb filtering effects with small head movements
- Stability of the above attributes when turning their heads and moving over a 1 meter radius area around the center

Listeners were also asked to describe their overall impressions.

Each listener was able to select between either of the two arrays with the three decoder configurations (six choices) by a single mouse click, so that it was unnecessary to move one's head while making comparisons. The file comparison program also allows for looping, which is useful for selecting a single small program segment to be compared in the various configurations.

The second group of listening tests compared the localization performance of the 1.732 : 1 rectangular array to that of a hexagonal array. This particular aspect ratio was chosen, in part, in the first tests because it could be extended to a hexagon by the simple expedient of adding two loudspeakers at the sides, and altering the patching during the switching. The loudspeakers at the side came within about 30 cm of the room boundaries, whereas the front and rear loudspeakers were at least 1 meter away from room boundaries. While it was desired to keep the loudspeakers away from room boundaries to avoid adding an additional variable, it was not possible to do this and simultaneously maintain the rectangular array for comparison. It should be noted that the problem of fitting the loudspeaker array into the listening room is one of the principal reasons for investigating elongated arrays such as the rectangular array.

4. DISCUSSION

In addition to the activities described earlier, listeners were asked to state their overall preference–if they could

choose one speaker array and decoder for use in their homes, which would it be?

The hexagonal array with shelf filter decoder was preferred above all other combinations when all the listening material was included in the test. When the listening material was limited to frontal source plus ambiance, the $1.7:1$ rectangular array with shelf filters was judged equal to the hexagonal array. Poor side imaging combined with a shift in tonal balance for sources directly ahead made the square array the least preferred of the three configurations tested.

Of the four decoder types tested–velocity, energy, shelf, cardioid–the shelf filter decoder was preferred for natural sources.

The cardioid and velocity decoders were least preferred. The velocity decoder produced comb-filtering aural artifacts when the subjects moved their heads as well as having the least stable side images. Some listeners also reported uncomfortable in-head or near-head imaging. This was more pronounced on recordings where the instruments were close to the microphone.

The cardioid decoder produced stable imaging with listener movement and no discernible combing artifacts, but test subjects felt that it was too diffuse, and too reverberant with natural sources. On the chamber orchestra recording, one of the listeners, who is familiar with the hall in which the recording was made, remarked that it did indeed sound like the hall, but the perspective was from much farther back in the hall, not the front where the microphone was placed.

The energy decoder provides a balance between these two extremes. Listeners judged that the reproduction was more focused and less diffuse than the cardioid decoder, without introducing obvious phase combing artifacts present in the velocity decoder.

The shelf filter decoder was preferred, as noted above because it dispensed with the in-head reproduction artifacts encountered with the velocity decoder, but retained a more focused image for individual sound sources. As an example, the recordings of the alto female voice sounded less spread out in space with the shelf filter decoder, as compared with the energy decoder.

We should note that the above are general impressions and more weight has been given to natural recordings than to the test signals. By and large, the test signals were used to help illuminate the differences between the different decoders and layouts. In almost all cases certain test subjects disagreed with the above rankings on specific source material.

Some listeners also noticed an accommodation effect where after prolonged listening to the test signals, the localization quality of all speaker arrays appeared to worsen. For example, side sources were more strongly drawn to the front speakers. This was especially noticeable with sibilant speech sounds. After a brief break in listening, the localization improved. This effect was more apparent when listening to the panned test signals than with natural sources. In particular, it was not observed with the applause and fireworks recordings, which were reproduced uniformly and without noticeable speaker-detent effect.

More testing is needed with a wider variety of sources–in particular studio recordings with sound source placed all around the listener. Such recordings were not available to the authors at the time the tests were performed. However, we feel that we can recommend the use of shelf filters for natural source material. Their use provides specific improvements in the focus and perspective of the reproduced audio, with no discernible negative artifacts. If shelf filters are not used, the energy decoder provides the best balance between spatial accuracy and audible artifacts with movement off the center. We also note that with material with a frontal emphasis, the rectangular speaker array performs better than the square.

5. CONCLUSIONS

- Of the six decoders (square, rectangle) $\times$ (velocity, energy, and shelf), all are noticeably different.
- The order of preference for decoders is for shelf filter, followed by the 'energy' decoder, followed by the 'velocity' decoder.
- The order of preference for loudspeaker layouts is hexagon, followed by rectangle, followed by square.
- Changes in layout make significantly more difference than changes in decoder.
- Neither the square array nor the rectangular array gives a satisfactory impression of images to the side. The hexagonal array does give a good impression of side images.

- For a square array, the localization quality of a sound source is different for front vs. front diagonal sound sources, despite the fact that the velocity and energy calculations show isotropic behavior.
- For some program material, shelf filters supply a focusing effect. This focusing effect has to do with bringing various spectral components into the same position.
- The test methodology works well, and something similar is necessary to gather meaningful opinions about differences between layouts or decoders.
- Sibilance is heard drawn to the front loudspeakers, to whichever side the source is on.
- Layout/decoder ranking is program material dependent.

The authors have found that the various decoder implementations recommended by Gerzon work well. The strong frontal localization performance of rectangular arrays has not been emphasized in previous publications and deserves attention, especially for rooms that cannot accommodate regular polygonal arrays.

6. FUTURE WORK

The test methodology used in this paper will be extended to testing larger loudspeaker arrays, loudspeaker arrays that are not regular, in the sense of having variable radius, and to testing reproduction arrays with height.

Experiments will be conducted to test the effect of perturbation of the exact array loudspeaker locations. How much error in speaker placement or speaker response can be tolerated?

7. WEB SITE

The authors have created a website at `http://www.ai.sri.com/ajh/ambisonics` where some of the the computer programs, test signals, and other material used in this work can be downloaded.

8. REFERENCES

[1] Michael A. Gerzon. General Metatheory of Auditory Localisation. In *Preprints from the 92nd Audio Engineering Society Convention, Vienna*, number 3306, March 1992. AES-EL 3366.PDF.

[2] Y. Makita. On the Directional Localisation of Sound in the Stereophonic Sound Field. *E.B.U. Review, Part A - Technical*, (73):102–108, June 1962.

[3] K. deBoer. Stereophonic Sound Production. *Philips Technical Review*, 5:107–144, 1940.

[4] Jens Blauert. *Spatial Hearing*. MIT Press, revised edition, 1996. ISBN: 0262024136.

[5] Michael Gerzon. Surround-Sound Psychoacoustics. *Wireless World*, 80(1468):483–486, December 1974. Available from `http://www.audiosignal.co.uk/Gerzon%20archive.html` (accessed June 1, 2006).

[6] Michael A. Gerzon. Multidirectional Sound Reproduction Systems. U.S. Patent 3,997,725, December 1976.

[7] Ken Farrar. Soundfield Microphone. *Wireless World*, 85(1526):48–50, October 1979. Part 2 in issue 1527, pp 99–103.

[8] Eric Benjamin and Thomas Chen. The Native B-Format Microphone. In *Preprints from the 119th Audio Engineering Society Convention, New York*, number 6621, October 2005. AES-EL 13348.PDF.

[9] Michael Gerzon. Multi-System Ambisonic Decoder. *Wireless World*, 83(1499):43–47, July 1977. Part 2 in issue 1500. Available from `http://www.geocities.com/ambinutter/Integrex.pdf` (accessed June 1, 2006).

[10] Michael A. Gerzon. Practical Periphony: The Reproduction of Full-Sphere Sound. In *Preprints from the 65th Audio Engineering Society Convention, London*, number 1571, February 1980. AES-EL 3794.PDF.

[11] Michael A. Gerzon and Geoffrey J. Barton. Ambisonic Decoders for HDTV. In *Preprints from the 92nd Convention of the Audio Engineering Society, Vienna*, number 3345, March 1992. AES E-lib `CD12/pp9193/pp9203/3405.pdf`.

[12] Bruce Wiggins, Iain Paterson-Stephens, Val Lowndes, and Stuart Berry. The Design and Optimisation of Surround Sound Decoders Using Heuristic Methods. In *Proceedings of UK-Sim 2003, Conference of the UK Simulation Society*, pages 106–114. UK Simulation Society,

2003. Avaialbe from http://sparg.derby.ac.uk/SPARG/PDFs/SPARG_UKSIM_Paper.pdf (accessed May 15, 2006).

[13] Octave. http://www.octave.org/ (accessed July 1, 2006).

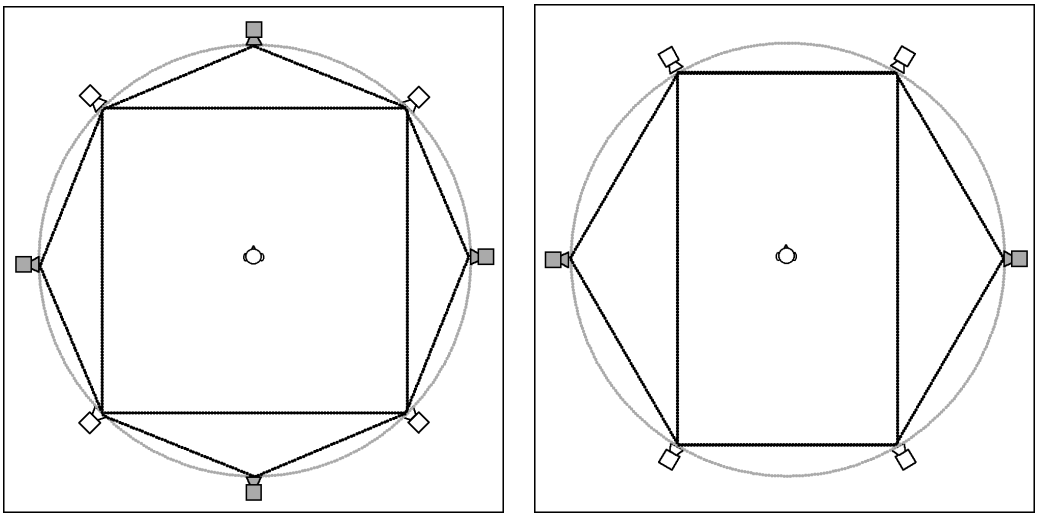
[14] Philip M. Morse and K. Uno Ingard. *Theoretical Acoustics*, chapter 7, page 311. Princeton University Press, 1986. ISBN: 0691024014.

[15] Richard Lee. Shelf Filters for Ambisonic Decoders. http://www.ambisonicbootlegs.net/Members/ricardo/SHELFs.doc/view (accessed July 3, 2006), 2005.

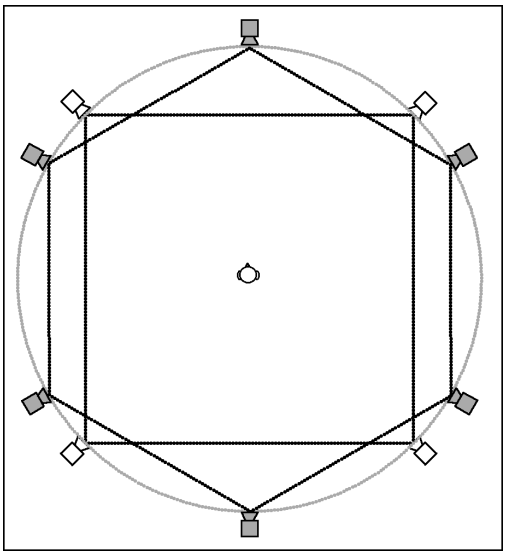
A.1 GNU OCTAVE CODE LISTING

```
%%
% GNU OCTAVE code to implement Fig 12 "The
% Design Mathematics" of M.A. Gerzon,
% "Practical Periphony: The Reproduction of
% Full-Sphere Sound" Preprint 1571 (A6)
% 65th Audio Engineering Society Convention,
% 2/1980, London
%%
%%  Aaron J. Heller <heller@ai.sri.com>
%%  Last update: 2006-08-01 14:05:42 -0700
%%

function retval = speaker_matrix(positions, k)
% SPEAKER_MATRIX - compute speaker decode matrix
%  positions are the XYZ positions of
%  the speaker pairs, one speaker pair per
%  row, i.e., [1 1 1; 1 -1 -1] If Z
%  positions of speaker pairs are omitted,
%  it does a horizontal decode (otherwise Z
%  gain is infinite).
%
%  Return values are the weights for X, Y,
%  and Z as columns of the matrix.
%
%  positions: [ x_1 y_1 z_1;
%                   ...
%               x_i y_i z_i;
%                   ...
%               x_n y_n z_n ]
%
%          k: 1         => velocity,
%             sqrt(1/2) => energy,
%             1/2       => controlled opposites
%
%  retval:  alpha_1 ... alpha_i  ... alpha_n
%           beta_1  ... beta_i   ... beta_i
%           gamma_1 ... gamma_i  ... gamma_n
%  where the signal for the i'th speaker
%  pair is:
%
%  S_i = W +/- ( alpha_i*X +
%                beta_i*Y +
%                gamma_i*Z )
%
%  Note: This assumes standard B format
%  definitions for W, X, Y, and Z, i.e., W
%  is sqrt(2) lower than X, Y, and Z.
%
%  Example:
%   octave> speaker_matrix( [1 1 ; 1 -1], 1 )
%   ans =
%
%      1.0000   1.0000
%      1.0000  -1.0000
%

  % allow entry of positions as
  % transpose for convenience
  positions = positions';

  % n = number of speaker pairs
  % m = number of dimensions,
  %     2=horizontal, 3=periphonic
  [m,n] = size(positions);

  % scatter matrix accumulator
  s = zeros(m,m);

  % speaker directions matrix
  directions = zeros(m,n);

  for i = 1 : n

    % get the i'th speaker position
    pos = positions(:,i);

    % normalize to get direction cosines
    dir = pos/sqrt(pos' * pos);

    % form scatter matrix and accumulate
    s += dir * dir';

    % form matrix of speaker directions
    directions(:,i) = dir;

  end

  retval = sqrt(1/2) * n * k * ...
           inverse( s ) * directions;

endfunction
```

A.2 ADDITIONAL SPEAKER LAYOUTS



(a) (b)



(c)

Figure A.1: Additional loudspeaker array comparisons. (a) square and octagonal (b) $\sqrt{3}:1$ rectangle and regular hexagon (c) square and regular hexagon. Comparisons of the first two arrays can be done with the current setup. The third will require additional D/A channels.