TOWARDS A RATIONAL BASIS FOR MULTICHANNEL MUSIC RECORDING

James A. Moorer, Sonic Solutions
Jack H. Vad, San Francisco Symphony

The DVD-Audio standard will include multi-channel uncompressed PCM audio. Although we have 30 (or more) years of experience with recording, mixing, and reproducing stereo, as an industry we have relatively little experience with multi-channel music recordings for the home. In this paper, we have chosen to optimize perceptual aspects of spatialization. This leads to a mathematical basis for pan matrices, microphone placement, and for speaker placement correction. Much of this basis can be considered to be an extension of the sound-field work pioneered by Michael Gerzon. Recordings were done using simultaneous, multiple experimental microphone techniques so that different methods could be easily compared. Results of these experiments are described.

Introduction:

The number of ways we could imagine to do live recordings for presentation in a multi-channel format is boundless. Although ultimately the final choices must be made on aesthetic grounds, it seems useful to try to develop some objective means of evaluating different techniques. The objective methods can then be used both to explain shortcomings of certain techniques and to lead us in different directions that we may not have considered otherwise.

We have chosen to optimize perceptual aspects of spatialization. This leads to a mathematical basis for pan matrices, microphone placement, and for speaker placement correction. There are two perceptual aspects that we consider: localization and ambiance. By localization, we mean the accuracy with which the direction of a sound source is perceived. By ambiance, we mean the feeling of "spaciousness" and depth. Although this may appear obvious, we note in passing that *the only possible contribution of multi-channel audio over stereo is in the spatial aspects of the sound.*

We must mention that music recordings are made for a wide variety of purposes. Different people will optimize different aspects of the sound. We cannot hope to second-guess the intent of the artists. The point of this examination is to attempt to explain on a rational basis what we are hearing with various recording and panning techniques and to help the artist achieve the desired effect, whatever that may be.

For localization, there are some objective measures that attempt to simulate human perception of direction. These were called by Gerzon the pressure vector and the energy vector. It is fairly well established that the pressure vector corresponds roughly to low-frequency direction perception, and the energy vector to high-frequency direction perception. In the mid-frequency range where our perception of direction is most acute, both vectors are important. We note a refinement to the energy vector calculation that brings it more into line with experimental determinations of loudness perception.

In the standard 5-channel surround setup that is popular with home theater surround-sound systems, the speakers are generally not located at equal angles. We develop techniques for dealing with varying placements of speakers, and techniques for re-matrixing the sound in the home for speaker placements that differ from the placements in the studio where the music was originally mixed. Subjective assessments of the efficacy of these techniques will be discussed.

An experimental recording was done during the regular concert series of the San Francisco Symphony Orchestra where several different microphone placements were simultaneously recorded on a multi-track recorder. This allows simple comparison of different techniques and different panning methods during the subsequent mixdown. Some subjective results will be presented.

This paper is an expanded version of one presented at the 103rd AES Convention in Fall of 1997 [1]. Some of the development will be repeated here.

Perceptual Theory of Direction

It is well established that human perception of direction is a relatively complex affair. We will simplify it in this exploration to a simple measure that handles two of the most important aspects of perception: inter-aural time delay and energy difference. These may be summarized in a single formula. We use the definition of angle and speaker number shown in Figure 1. Note that this definition of angle differs from the standard mathematical definition. The center of the coordinate system is at the center of the listener's head.

Although it might be nice to have speakers out of the plane for full 3-dimensional ("periphonic") presentation [2], this is not likely to be the common setup in the home. For this purely practical reason, we will restrict ourselves to planar speaker arrays.

Gerzon [3] presented a summary of some aspects of directional theory that included calculation of the Makita (velocity) localization vector and the power vector. These correspond roughly to first- and second-order aspects of localization. As Gerzon points out, our knowledge of psychoacoustics suggests that human perception of location at low frequencies is dominated by velocity cues and by power cues at high frequency [4]. We will define a single measure that includes both of these as special cases.

Let g_i be the gain of the signal fed to speaker *i,* at an angle of θ_i.

(1) $$\theta_\gamma = \tan^{-1}\left(\frac{\sum g_i^\gamma \sin\theta_i}{\sum g_i^\gamma \cos\theta_i}\right)$$

For low-frequency direction, Bernfeldt [5] showed that inter-aural time-delay is equivalent to (1) with $\gamma = 1$. We will call this the velocity, or pressure vector. Gerzon used an energy vector calculation that is equivalent to (1) with $\gamma = 2$. An exponent of 2 is very convenient for mathematical manipulation, and it makes it possible to derive explicit closed-form results, such as the "Rectangle Decoder Theorem" [3, 7]. Unfortunately, it does not correspond exactly to the perception of direction at high frequencies. For instance, the ear's power law does not suggest an exponent of 2, but of some smaller value. Using Stevens' result that a 10 dB change in signal power results in a perceptual doubling of loudness [6], we might suggest that $\gamma = 1.66$ would be a more suitable approximation. This still does not take "head shadow" effects into account. A more complete objective measure would simulate the effects of head shadowing as well. We will not attempt to do that here, but will use the simple approximation used by Gerzon of $\gamma = 2$. The results agree roughly with subjective evaluations, so this seems to be a reasonable starting point.

Gerzon's "Rectangle Decoder Theorem" shows that for speakers in a perfect square, these vectors can be made to correspond when the channels are fed by three independent signals. If we let *W, X,* and *Y* represent these 3 independent signals, then we present the speakers with the following signals:

(2) $$S_i = W + Y\cos\theta_i + X\sin\theta_i$$

Again, θ_i is the angle of speaker *i* using the coordinate system shown in Figure 1. S_i is the signal that goes to speaker *i. W* is the signal that would be picked up by an omnidirectional microphone at the origin of the coordinate system. *Y* is the signal that would be picked up by a figure-of-eight pattern microphone pointed forward (towards angle zero). *X* is the signal that would be picked up by a figure-of-eight microphone pointed to the left (towards 90°)[1].

[1] Note that this microphone placement is for the purpose of mathematical examination of directional aspects of sound. It is not suggested here as a practical microphone technique. Among other things, the

It is straightforward to show that with the speaker signals shown in equation (2), using 4 speakers in a perfect square, the velocity and energy vectors correspond exactly. Gerzon also generalized the theorem to any regular polygon [7]. Any number of speakers (greater than or equal to 4) can be placed at equal angles and the vectors will still correspond.

There are a few corollaries that can be derived from Gerzon's proof of the rectangle decoder theorem. They are as follows:

- Uniqueness: *Only speaker signals of the form shown in equation (2) will allow the velocity and energy vectors to align.*
- Specificity: *If the speaker placement is not equi-angular (i.e., not in a regular polygon), then there is no drive that will cause the vectors to align for all angles.*[2]

For non-equi-angular placements, such as that shown in Figure 1, the velocity and energy vectors can not, in general, be made to align everywhere. The difference between these vectors can be used as a measure of the perceptual "spreading" of the image. The difference between the "true" angle of presentation and either of these calculated angles can be used as a measure of the accuracy of the reproduction of angle.

We will subsequently develop a technique for minimizing the difference between the vectors. Appendix B shows the development of a closed-form solution for the speaker gains that produce the minimum difference between the vectors for irregular speaker placement. The vectors can be made to match for most angles, but not for all.

Placement of a Sound: Pan Matrices:

With stereo sound, there is not much choice in how to distribute the sound to two speakers: you send some amount to the left and some amount to the right. We can debate endlessly what the correct proportions are, but we still only have 2 degrees of freedom (or one if overall loudness is controlled separately). With multiple speakers, the question is a bit more subtle. For instance, we might try the "obvious" method, which could be described as follows: to place a sound at a particular angle, you find the two closest speakers, then apply stereo panning to those two speakers to locate the sound in between them. We can apply (1) to this technique and plot the results. Figures 3 and 5 show the gains to the 5 speakers using stereo panning techniques to adjacent speakers. Figure 3 shows a -6 dB crossover point, and Figure 5 shows a -3 dB crossover point. For these purposes, the speaker placement is assumed to be symmetric about the front-to-back axis, so we have limited these plots to 180°. Figures 4 and 6 show the pressure and energy angles determined by this technique. These angles are the deviations from the intended angle, and thus can be considered to contribute to the perceptual error. The pressure angle uniformly shows less error than the energy angle. The difference between these curves represents the "spreading" of the image.

---

frequency responses of directional microphones, such as figure-8 patterns, differ significantly at low-frequencies from that of the omnidirectional microphone that these can not be used without equalization.

[2] The proof of the decoder theorem for regular polygonal speaker placement relies on the fact that

$$\sum_{i=0}^{N-1} \cos \frac{2\pi i}{N} = 0$$

If the speakers are not equi-angular, then cancellation of the sum of cosines of speaker positions does not occur, and the vectors fail to coincide.

In all the plots, we assume speaker angles of 60° and 135°. This is a somewhat wider placement of the front speakers that can be achieved in most homes. We will discus later the implications of moving the front speakers closer together.

The first thing that should be apparent is that neither angle is correct. The perceived location will deviate from the desired location except when the direction is either directly into one of the speakers, or directly in between two speakers. There is significant deviation at all other angles. Moreover, the preferred technique, which has a -3 dB crossover point, shows significant image spreading, since the energy angle error is in the opposite direction from the pressure angle error. To interpret these plots, we must remember that a positive deviation means that the image is "pulled" towards the next speaker, and a negative deviation means that the image is "pulled" towards the previous speaker.

Note that since we are talking about a pan matrix, we could simply re-label the pan knob with the angle that we perceive the sound at. This would straighten out the curves somewhat, but the important fact is the difference between the two curves, since this represents the diffusion of the image. Additionally, this technique would not work at all for positioning spot microphone feeds into a multi-channel recording using the 2-dimensional sound-field microphone which is described later.

Using a spatial harmonic expansion, we will derive a pan matrix that exhibits perfect pressure angle. The energy angle will then be a measure of the image spreading involved.

We will start with the Fourier sine and cosine series on the circle. This is equivalent to the spin harmonic functions described in [2], but reduced to the 2-dimensional case. In this representation, the directivity function of a sound at a certain angle ϕ can be expressed as follows[3]:

(3) $$\frac{1}{2}+\sum_{n}\cos n(\theta-\phi)$$

This represents a "spatial impulse" in the direction ϕ. The contribution to directivity from speaker *i* is then:

(4) $$g_i\left\{\frac{1}{2}+\sum_{n}\cos n(\theta-\theta_i)\right\}$$

where g_i is the gain of the signal going to that speaker. This is simply the Fourier sine and cosine series for a sound originating in the direction of speaker *i*;

We may now calculate the unknown channel gains, g_i, by fitting the directivity function (equation (3)) to the directivity function obtained by our speaker placement (equation (4)). This fit may be performed as a least-squares operation to determine the unknown channel gains[4]. After some manipulation, we arrive at the set of linear equations as follows:

[3] This formula is derived in Appendix A of reference [1] and will not be derived again here.

[4] Just to make it explicit, subtract (3) from (4), square it, and integrate with respect to θ over 2π, then differentiate with respect to the unknowns and set the result to zero.

(5) $$\sum_{i=1}^{N} g_i \left\{ 1 + 2\sum_n \cos n(\theta_i - \theta_k) \right\} = 1 + 2\sum_n \cos n(\phi - \theta_k)$$

for $k = 1,2,...,N$. This gives N equations in the N unknown channel gains. In our case, of course, $N = 5$. To be perfectly clear, both θ_i and θ_k are the angles of the 5 speakers using the coordinate system shown in Figure 1.

The astute reader will notice that the bounds on the summation on the cosine series have been routinely omitted. This is deliberate. Technically, the expansion is not limited. Since it is non-convergent, this is somewhat unhelpful. It does have to be bounded. In fact, given that we only have 5 speakers, sampling theory says that if they were spaced equi-angularly (at the vertices of a regular N-gon), then at best, we could recreate only the first two terms ($n=1$ and $n=2$). Since our speakers are not equi-angular, it is spatial sampling with unequal steps. The most conservative reading of the sampling theorem dictates that the highest spatial harmonic is then related to the largest step. This limits us for practicality to the first term only. In fact, if any of the angles between successive speakers is greater than 90°, then even the first spatial harmonic can not be recreated exactly, and there is no hope for the second harmonic - it will be aliased.

Before we examine the implications of this equation, it is worth discussing the higher harmonics briefly. If we extend the series up to any given term then stop, we will get the spatial analog of the Gibbs phenomenon. That is, there will be ripple, side-lobes, and other undesirable side-effects. In general, the series would have to be limited by applying a window function. This can assure a smooth directionality function, at the expense of widening the image somewhat. This is a well-known tradeoff. Unfortunately, this implies that you need a number of terms. A reasonable window function is not possible without 5 or 6 terms of the series, which would imply 11 to 13 speakers. Anything between 5 speakers and 11 speakers can not make effective use of higher spatial harmonics, because of the side effects of abrupt termination of the spatial harmonic expansion. About the best we can do is to use the 0^{th} and 1^{st} harmonics, with possibly some small amount of 2^{nd} harmonic to achieve a certain effect. We will discus this further a bit later. As an example of the effect of un-windowed truncation, Figure 2 shows one solution of Equation (5) for the front speaker using the 0^{th}, 1^{st} and 2^{nd} harmonics. Note that it is bimodal: as the angle of the sound goes from 0° (directly in front of the listener) to 180° (directly behind the listener), the gain goes to a minimum, but then rises to a secondary maximum. If a window function is applied (*i.e.*, only a fraction of the 2^{nd} harmonic is used), this behavior can be controlled.

There are a few properties of solutions to this equation that fall out immediately. The first is that the sum of all the speaker gains is unity. This may or may not be the desired effect. Many people prefer that the sums of the *squares* of the speaker gains be unity. Obviously, the gains can be renormalized so that any particular property is satisfied. We might suggest that it would be better to require equal *loudness*, which is a perceptual measure [8]. Unfortunately, the loudness is a relatively difficult quantity to compute, since it involves a frequency analysis of the sounds that are being presented so that masking phenomena can be taken into account. Given that we wish to calculate the gains of a pan matrix without reference to the sound that will be presented, we might normalize the gains by the following factor:

$$(6) \qquad \frac{1}{\sqrt[\gamma]{\sum_i g_i^\gamma}}$$

We recognize this as the γ-norm of the speaker gains. As noted above, the value $\gamma = 1.66$ gives an approximation to equal loudness at any desired angle in a frequency-independent manner.

Another property of any solution to equation (5) is that the pressure angle is perfect. Figures 7 and 8 show the gains and the resulting angles from a solution to equation (5). The horizontal line in Figure 8 shows that there is no error in the pressure angle.

Equation (5) is of rank 3, since there are only 3 independent variables - one 0^{th}-harmonic term and two 1^{st}-harmonic terms. To get a unique solution, we need to add other constraints. Probably the most straightforward extension is to require that the second harmonic terms be specified. This gives us 5 independent equations in 5 unknowns. Figures 7 and 8 were produced from requiring the second harmonic terms be zero. In fact, the second harmonic term corresponding to the cosine term should, in general, be set to zero if we do not have 11 or more speakers. The reason is that a non-zero second-harmonic cosine term produces asymmetry in the speaker feeds. That is, the gains for a particular angle will not be the mirror-images of the gains for the negative of the angle. This is weird enough to warrant elimination from further consideration. The sine term, however, can be useful and will be carried along further.

This system of equations can be solved in closed-form. Appendix A gives the formulas. They are complex enough that it is probably easier to solve the system of equations rather than use the closed-form solution, since the system of equations is quite well-conditioned. The solution does point out some degenerate conditions that probably should be obvious - such as when two speakers are placed at exactly the same angle. Note also that when the speakers are located at equal angles, most of the terms drop out.

With a constant 2^{nd}-harmonic, the curves shown in Figure 7 are sinusoids.

There is one striking difference between the solution for channel gains given by equation (5) and what you might imagine for surround-sound pan pots, and that there are contributions from *all* the speakers, and sometimes speakers are driven out-of-phase. This is somewhat non-intuitive but it is required to preserve the spatial harmonics.

Notice that the left side of equation (5) depends only on the speaker placement. This matrix can be computed and inverted once for a given speaker layout. For any desired angle, the channel gains may then be computed by a simple matrix multiplication. We will be guaranteed that the zero-th and first spatial harmonics will agree with the zero-th and first spatial harmonic of a sound source at the given angle.

The second harmonic term can be set to any constant value without changing these properties. Too much second harmonic produces undesirable effects, such as certain channel gains being entirely negative.

Since the entire system is linear, and we generally assume that air is linear as well, we may then add voice after voice, each at its own angle, using equation (5) to determine the channel gains.

Note that the one "free" variable is the constant second-harmonic term. Since this may be set to any value, we may use it to optimize any aspect we choose. Specifically, it may be used to minimize the difference between the pressure and energy vectors. If we choose $\gamma = 2$ in equation (1), then there is a closed-form solution shown in Appendix B. For $\gamma = 1.66$, or any other value, the solution can not be expressed in closed-form, but can easily be found by use of a line search [9]. Unfortunately, the solution is not unique, and even the closed-form solution is not particularly well-behaved. For most of the desired directions, the solution brings the energy vector and the pressure vector into exact coincidence. This fact is interesting, but not terribly useful because of the poor behavior of the solution. The optimum value of second harmonic sometimes requires large gains with large negative gains in the opposing speakers. This can produce phasing for listeners that are not at the exact center of the speakers. For some angles, the optimum second harmonic can tighten the spatial image a bit, but for other angles, it produces unacceptable solutions.

Correction for Speaker Placement:

There is an interesting additional use for equation (5), and that is for adapting a recording to a different set of speaker positions. If a recording is mastered using a given setup with speaker angles θ_i, we may compute a matrix that converts the original speaker signals into another set of signals such that the zero-th and first spatial harmonics correspond.
To see this, define the matrix representing the right-hand side of equation (5) as follows:

(7) $$M(k,i) = 1 + 2\sum_n \cos n(\theta_i - \theta_k)$$

For 0th and 1st order harmonics, we need to augment this equation with two different rows:

(8) $$M(4,i) = \sin(2\theta_i)$$

(9) $$M(5,i) = \cos(2\theta_i)$$

Similarly, we define the speaker gain vector as follows:

(10) $$G = [g_1, g_2, g_3, g_4, g_5]^T$$

If we use the overbar to represent the matrix and new gains using the new speaker placement, then we can solve for the new speaker gains as follows (in matrix notation):

(11) $$\overline{G} = \overline{M} M^{-1} G$$

Note that aesthetically, this may or may not be what is desired. The artist may want a particular sound to come out of, say, the left front speaker regardless of where the speaker is located in the room. In the case where a particular angle is desired, regardless of speaker placement, the above procedure will accomplish that goal. This is also appropriate when using the 2-dimensional sound field microphone, as described in the next section.

We might point out here that equation (11) can also be used to rotate the sound field. Since we can use any set of speaker position angles to form the matrix $\overline{M}$, we could, for instance, place the “center” speaker on the left, or even behind us. This is of dubious utility in the home.

Application to Music Recording:

We may use the same theoretical basis to derive a recording method that can be presented in any number of channels that preserves the angular position of sounds at the time of recording. Figure 9 shows a 2-dimensional sound-field microphone that consists of 3 identical directional capsules. A particularly interesting arrangement is to use 3 hypercardiod capsules and separate the forward-facing capsules, as in the ORTF arrangement. This simultaneously gives us a stereo recording, using the front-facing capsules, and a multi-channel recording when the effects of the rear-facing capsule is added.

The directional feeds from the microphones may be described as follows:

(12) $m_1(\theta) = C + (1 - C)\cos(\theta - \Omega)$

(13) $m_2(\theta) = C - (1 - C)\cos(\theta)$

(14) $m_3(\theta) = C + (1 - C)\cos(\theta + \Omega)$

Where C represents the directionality of the capsule. For a hypercardoid capsule, $C = \frac{1}{4}$. Of course, θ represents the angle of the sound source using the coordinate system shown in Figure 1, and Ω represents the angle of the axis of the forward-facing capsules, as shown in Figure 9. For the ORTF arrangement, this would be 54.7356°.

From these three feeds, we can derive the 0^{th} and 1^{st} harmonic components as follows:

(15) $$\frac{a_0}{2} = \frac{\frac{1}{2}(m_1 + m_3) + m_2\cos(\Omega)}{C(1 + \cos(\Omega))}$$

(16) $$a_1 = \frac{\frac{1}{2}(m_1 + m_3) - m_2}{(1 - C)(1 + \cos(\Omega))}$$

(17) $$b_1 = \frac{m_1 - m_3}{2(1 - C)\sin(\Omega)}$$

where a_0 represents the coefficient of the 0^{th}-order spatial harmonic, and a_1 and b_1 represent the coefficients of the cosine and sine terms of the 1^{st}-order spatial harmonic.

We can then derive the equations to matrix each microphone feed into the signals to deliver to the speakers. Let S_i be the signal fed to speaker *i*. Let the column vector *S* represent the speaker feeds (a 5-by-1 matrix). Let *R* be a 5-by-1 matrix representing the right-hand side of equation (5). We can then write *R* explicitly as follows:

$$(18) \qquad R = \begin{bmatrix} a_0 + 2a_1\cos(\theta_1) + 2b_1\sin(\theta_1) \\ a_0 + 2a_1\cos(\theta_2) + 2b_1\sin(\theta_2) \\ a_0 + 2a_1\cos(\theta_3) + 2b_1\sin(\theta_3) \\ 0 \\ 0 \end{bmatrix}$$

The coefficients a_0, a_1, and b_1 are combinations of the three microphone feeds as noted in equations (15), (16), and (17). We can then solve explicitly for the speaker feeds as follows:

$$(19) \qquad S = M^{-1}R$$

where *M* is the matrix defined by equations (7), (8) and (9). This can be expanded further to make the contribution of the microphones explicit. We will factor *R* as follows:

$$(20) \qquad R = R_1 R_2 \begin{bmatrix} m_1 \\ m_2 \\ m_3 \end{bmatrix}$$

$$(21) \qquad R_1 = \begin{bmatrix} 1 & \cos(\theta_1) & \sin(\theta_1) \\ 1 & \cos(\theta_2) & \sin(\theta_2) \\ 1 & \cos(\theta_3) & \sin(\theta_3) \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$(22) \qquad R_2 = \begin{bmatrix} \frac{1}{2C(1+\cos(\Omega))} & \frac{\cos(\Omega)}{C(1+\cos(\Omega))} & \frac{1}{2C(1+\cos(\Omega))} \\ \frac{1}{2(1-C)(1+\cos(\Omega))} & \frac{-1}{(1-C)(1+\cos(\Omega))} & \frac{1}{2(1-C)(1+\cos(\Omega))} \\ \frac{1}{2(1-C)\sin(\Omega)} & 0 & \frac{-1}{2(1-C)\sin(\Omega)} \end{bmatrix}$$

Thus the speaker feeds can be expressed as follows:

$$(23) \qquad S = M^{-1}R_1 R_2 \begin{bmatrix} m_1 \\ m_2 \\ m_3 \end{bmatrix}$$

Figure 10 shows the gain of the rear-facing microphone into the center speaker, and the gains of the forward-facing microphones into the center speaker versus the angle of speakers 2 and 5 (front left and front right). Note the interesting result that when the angle of the speakers reaches 45°, the gains to the center speaker go to zero. At lower angles, the center gain goes negative. This tells us several things:

- For presentation of music, the front speakers should be spaced further apart than is generally done for film sound presentation.
- At “reasonable” speaker angles, the feed to the center speaker is small, which should increase the decorrelation of sound from the left versus sound from the right.
- If the front speakers are too close together, the center speaker feed becomes large and negative. This is to be avoided due to phasing problems when the listener is not in the exact center of the speakers.

We may then analyze this method of distributing the microphone feeds to the 5 channels by applying equation (1) to determine the pressure and energy angles. They will, of course, be exactly those shown in Figure 8. That is, *the pressure angle will correspond exactly to the angle of the original sound.* Furthermore, the energy angle will only deviate by no more then 5° in front and no more than 6° in the rear.

Note also that equation (11) may be used to correct the signals to the 5 speakers to a different speaker layout. That is, we can master the recording using the geometry of the studio monitor system, then by a simple matrix multiplication, we can produce 5 speaker feeds for the home system such that the sounds will appear at exactly the same angles as they did in the studio. As noted before, this may or may not be what is desired.

Analysis of the Multiple Omni Technique

Note that there is another microphone technique that has been suggested for multi-channel recording. This is to hang omnidirectional microphones such that the placement of the microphones corresponds to the speaker locations that are used for auditioning. In this arrangement, there is one microphone for each speaker. There is no matrixing. Each microphone feed is sent to exactly one speaker. For an orchestral recording, this generally is done as 3 microphones placed over the orchestra and two more placed over the audience. If we analyze this technique, the resulting angles are shown in Figure 11. The levels to each speaker are shown in Figure 12. From this, we may immediately make the following predictions:

- *The angles are only accurate for a sound straight ahead or in back*
- *The image will be "drawn" towards the speaker location. There will be little localization of sounds between the speakers.*
- *The great difference between the pressure and energy vectors show that the imaging will be diffuse.*

The argument that is usually made for this arrangement is two-fold: first, since omnidirectional microphones have flat frequency responses, there is no coloration of the sound due to the directionality of the microphone. Second, since there is no matrixing, the sound coming from the left is very different from the sound coming from the right, so the left-right decorrelation should be large. There is, of course, feed-through of the sounds on the right to the microphones on the left, and vice-versa. These crossover sounds will be delayed and attenuated by an amount that can be calculated from the $1/d^2$ law, where *d* is the distance from the sound source. For a sound directly in front of the listener, the center speaker will receive the greatest signal, but it will also appear in the left and right speakers, though at a somewhat lower amplitude.

An Experimental Recording

To test the theories that have been presented, we did a recording during the regular concert season of the San Francisco Symphony in Louise M. Davies Symphony Hall. A total of 20 microphones were used to simultaneously test several different philosophies of microphone placement. These were as follows:

- One sound-field microphone as shown in Figure (9) located above the 3rd row of the audience using hypercardoid capsules.
- One ORTF pair to the left and right of the sound-field microphone.
- Two omnidirectional microphones 15' apart above the 8th row of the audience
- Four omnidirectional microphones suspended over the front of the stage about 8' apart.
- Nine "spot" microphones placed over various specific instrument groups (3 woodwind, 2 brass, violins II, viola, harp, and tympani).

The locations of the microphones was measured, so all the angles and distances were known. A schematic diagram of the microphone placement is shown in Figure 13.

The combinations of the omnidirectional microphones allowed us to test the idea of sending one microphone to one speaker. Since we were unable to place one single microphone over the conductor's head because of a permanently fixed speaker cluster, we used two microphones on each side of the speaker cluster for the center feed.

We recorded two different sound-field examples: one entirely coincident, sometimes called a "point" microphone, and one with the front-facing capsules separated by 17 centimeters. A single rear-facing hypercardoid was used for the rear feed.

A number of "spot" microphones were also used so we could test different ways of mixing the spot feeds into the surround matrix.

One of the pieces, Bartok's "Concerto for Orchestra", proved an excellent subject for these tests, since it features various combinations of tone colors, including solos, small ensembles, and full orchestra, with the ensembles using a variety of different combinations of instruments.

Experimental Results

Unfortunately, any reporting of the results of the experiment are necessarily subjective. Regardless of what our opinions are as to the relative quality of the different techniques, there will surely be differing opinions throughout the industry. As mentioned earlier, different artists will use different microphone placement and surround matrices to achieve differing aesthetic results. With this in mind, let us present *our* opinions as to the results of the test:

The omnidirectional pickup, where each microphone goes to exactly one speaker (except, as mentioned, the sum of the two center microphones goes to the center speaker) produced roughly the results predicted by the equations and shown in Figure 11. There was very little imaging between the speakers. One could clearly hear the speaker locations. There was little feeling of "space" in the concert hall, despite the fact that the reverberation in the hall was quite evident in the recording.

This result is somewhat curious. Two omnidirectional microphones, possibly supplemented with spot microphones, is a commonly used technique for stereo recordings. Indeed, if we limit the presentation to two speakers, we get somewhat more imaging between the speakers. Instrumental groups still get "pulled" towards the nearest speaker, but there is some feeling of space. Going from 2 speakers to 5 speakers did not make things better. In our opinion, it made things worse.

The coincident sound-field microphone and the ORTF sound-field microphones were quite similar with the ORTF front-facing capsules giving slightly better separation for some instruments. We tried switching between the stereo recording, where the two front-facing capsules were directed to left and right speakers, versus the 5-channel matrix of equation (19). The 5-channel presentation was superior in all respects, but the difference between the stereo presentation and the 5-channel presentation was not overwhelming. Adding the rear-facing microphone does definitely add a feeling of space and of being "immersed" in the sound. The imaging is improved, but (in our opinion) it still falls short of the experience of actually being in the concert hall.

One problem that became immediately evident was that our sound-field microphone was placed too far back. The audience sounds were clearly evident in the front speakers. The array should probably be over the conductor's head.

The use of the spot microphones is somewhat problematic. They do invariably pick up some sound other than the intended local group of instruments. We panned the spot microphones to the original angle of the microphone using equation (5) to determine the speaker gains. This preserved the imaging of the particular spot quite well, but it had a tendency to diffuse the imaging of instruments to the left and right of the spot microphone. This, presumably, is the tradeoff we must make when using spot microphones.

Summary:

We have presented a method of analyzing multi-channel pan matrices and microphone placement techniques for imaging quality in a manner that is largely consistent with human perception and corresponds to subjective evaluations. This technique was used to demonstrate that a straightforward extension to stereo panning formulas to multi-channel usage does not produce good imaging except when the sound is placed in one speaker only, or is placed halfway between two adjacent speakers. We also demonstrated that the method of making multi-channel recordings that consists of using a number of omnidirectional microphones then assigning each microphone to one and only one speaker with no matrixing can not produce good imaging.

The method of spatial harmonics was used to derive multi-channel pan matrices that produce reasonable imaging. This method was also used to suggest a microphone technique. The 2-dimensional sound-field microphone shown in Figure (9) has one significant advantage, in that the front-facing microphones can be used directly for stereo recording as well as multi-channel recording. The equations for matrixing the microphones into the speakers were given. Spot microphones may also be used to enhance certain performers or instrumental groups, at the expense of diffusing the imaging of surrounding instruments somewhat. Panning the spot microphone feed to the angle of the microphone by using the same matrix that was used for the sound-field microphone assures that the image from the spot microphone is in exactly the same position as in the sound-field microphone.

Conclusion:

Recording for multi-channel presentation of music requires changes to each stage of the process, from microphone placement to pan matrices. Since we will be making dual releases for some time to come (current stereo CD format and multi-channel DVD-Audio format), it seems important to make sure that our recordings are compatible with both release formats. We have given methods that have been verified experimentally (albeit subjectively) that can accomplish this.

Future Directions:

This study suggests as many issues as it answers. It would be interesting to move the spot microphones closer to the instruments so that there is less “leakage” from nearby instruments. This would sharpen the imaging, since the leakage diffuses the imaging of the surrounding instruments. There are, of course, a number of problems with “close” microphone placement, including, among other things, the pick up of extraneous sounds, and the sometimes unnatural sound arising from the directional characteristics of the sound field of the instrument itself.

Listening to the rear-facing microphone by itself is quite interesting. It suggests that it might be possible to “synthesize” the rear channel. This would be advantageous, since it would eliminate one prominent source of crowd noise. There are some interesting technical challenges, since it would involve trying to estimate an impulse response that is more than 100,000 points long. Additionally, the front-facing microphones do pick up some amount of the rear signal, so any synthesized rear signal, especially if it used a different impulse response (*i.e.*, it came from a different concert hall), would not correlate properly with the front-facing microphones.

Acknowledgment:

We would like to extend our gratitude to the San Francisco Symphony for their generous assistance in this project.

References:

[1] Moorer, James A. "Music Recording in the Age of Multi-Channel" , presented at the 103rd Convention of the AES, September 1997, New York, preprint #4623

[2] Gerzon, Michael A. "Periphony: With-Height Sound Reproduction" J. Audio Eng. Soc., Vol. 21, No. 1, Jan/Feb 1973, pp2-10

[3] Gerzon, Michael A. "The Optimum Choice of Surround Sound Encoding Specification" presented at the 56th AES Convention, March 1-4, 1977, Paris, France, Preprint number 1199 (session A-5).

[4] Durlach, Nathaniel I., and Colburn, H. Steven "Binaural Phenomena", Chapter 10 in "Handbook of Perception: Volume IV - Hearing" Edward C. Carterette, Morton P. Friedman, eds, Academic Press, New York, 1978, pp365-406

[5] Bernfeldt, Benjamin "Simple Equations for Multi-Channel Stereophonic Sound Localization," presented at the 50th Convention of the AES, 1975, London

[6] Stevens, S.S. "The Measurement of Loudness," Journal of the Acoustical Society of America, Volume 27, #5, Spetember 1955, pp815-829

[7] Gerzon, Michael A. "The Rational Systematic Design of Surround-Sound Recording and Reproduction Systems" unpublished manuscript, circa 1975.

[8] Zwicker, Eberhard, and Scharf, Bertram "A Model of Loudness Summation," Psychological Review, Volume 72, Number 1, 1965, pp3-26

[9] Fletcher, Roger "Practical Methods of Optimization," John Wiley & Sons, New York, 1989

Appendix A: Closed-form Solution to Pan Gains:

The closed-form solution can be expressed as follows:

(A1) $$g_\alpha = g_{\alpha\alpha+1} r_{\alpha+1} + g_{\alpha\alpha} r_\alpha + g_{\alpha\alpha-1} r_{\alpha-1} + g_{\alpha\sigma}\sigma$$

Where g_α is the gain to speaker α and σ is the (constant) 2nd harmonic contribution (which may be set to zero). r_α represents the right-hand side of the matrix equation. In the case of the pan matrix, r_α is set as follows:

(A2) $$r_\alpha = 1 + 2\cos(\theta_\alpha)$$

In the case of microphone feed, r_α is set as follows:

(A3) $$r_\alpha = a_0 + 2a_1\cos(\theta_\alpha) + 2b_1\sin(\theta_\alpha)$$

where a_0, a_1, and b_1 are set as shown in equations (15), (16), and (17).

We may now define $g_{\alpha\alpha-1}$, $g_{\alpha\alpha}$, $g_{\alpha\alpha+1}$, and $g_{\alpha\sigma}$ as follows, using μ, ρ, and τ as auxiliary variables.

(A4) $$\mu = 2\sin\frac{1}{2}(\theta_\alpha - \theta_{\alpha-2})\sin\frac{1}{2}(\theta_{\alpha+2} - \theta_\alpha)\left(-\sin(\theta_{\alpha+1} - \theta_\alpha) + \sin(\theta_{\alpha+1} - \theta_{\alpha-1}) - \sin(\theta_\alpha - \theta_{\alpha-1})\right)^2$$

(A5) $$\begin{aligned} g_{\alpha\alpha-1} = &-\frac{1}{\mu}\sin\frac{1}{2}(\theta_{\alpha+1} - \theta_\alpha)\sin\frac{1}{2}(\theta_{\alpha+1} - \theta_{\alpha-1}) \\ &(2\cos\frac{1}{2}(-\theta_{\alpha+2} - \theta_\alpha + \theta_{\alpha-1} + \theta_{\alpha-2}) + 2\cos\frac{1}{2}(-\theta_{\alpha+2} + \theta_\alpha - \theta_{\alpha-1} + \theta_{\alpha-2}) \\ &+ 2\cos\frac{1}{2}(\theta_{\alpha+2} - \theta_\alpha - \theta_{\alpha-1} + \theta_{\alpha-2}) + \cos\frac{1}{2}(-\theta_{\alpha+2} - 2\theta_{\alpha+1} + \theta_\alpha + \theta_{\alpha-1} - \theta_{\alpha-2}) \\ &+ \cos\frac{1}{2}(\theta_{\alpha+2} - 2\theta_{\alpha+1} - \theta_\alpha + \theta_{\alpha-1} + \theta_{\alpha-2}) + \cos\frac{1}{2}(\theta_{\alpha+2} - 2\theta_{\alpha+1} - \theta_\alpha + \theta_{\alpha-1} + \theta_{\alpha-2}) \\ &) \end{aligned}$$

(A6) $$\rho = 32\sin\frac{1}{2}(\theta_\alpha - \theta_{\alpha-2})\sin\frac{1}{2}(\theta_{\alpha+2} - \theta_\alpha)\sin^2\frac{1}{2}(\theta_\alpha - \theta_{\alpha-1})\sin^2\frac{1}{2}(\theta_{\alpha+1} - \theta_\alpha)$$

(A7)

$$g_{\alpha\alpha} = \frac{1}{\rho}(4\cos\frac{1}{2}(\theta_{\alpha+2} - \theta_{\alpha-2})$$

$$+ 2\cos\frac{1}{2}(\theta_{\alpha+2} - 2\theta_{\alpha-1} + \theta_{\alpha-2}) + 2\cos\frac{1}{2}(\theta_{\alpha+2} - 2\theta_{\alpha+1} + \theta_{\alpha-2})$$

$$\cos\frac{1}{2}(\theta_{\alpha+2} + 2\theta_{\alpha+1} - 2\theta_{\alpha-1} - \theta_{\alpha-2}) + \cos\frac{1}{2}(\theta_{\alpha+2} - 2\theta_{\alpha+1} + 2\theta_{\alpha-1} - \theta_{\alpha-2})$$

$$)$$

(A8)

$$g_{\alpha\alpha+1} = \frac{1}{\mu}\sin\frac{1}{2}(\theta_{\alpha} - \theta_{\alpha-1})\sin\frac{1}{2}(\theta_{\alpha+1} - \theta_{\alpha-1})$$

$$(2\cos\frac{1}{2}(-\theta_{\alpha+2} - \theta_{\alpha+1} + \theta_{\alpha} + \theta_{\alpha-2}) + 2\cos\frac{1}{2}(-\theta_{\alpha+2} + \theta_{\alpha+1} - \theta_{\alpha} + \theta_{\alpha-2})$$

$$+ 2\cos\frac{1}{2}(\theta_{\alpha+2} - \theta_{\alpha+1} - \theta_{\alpha} + \theta_{\alpha-2}) + \cos\frac{1}{2}(-\theta_{\alpha+2} - \theta_{\alpha+1} + \theta_{\alpha} - 2\theta_{\alpha-1} + \theta_{\alpha-2})$$

$$+ \cos\frac{1}{2}(\theta_{\alpha+2} - \theta_{\alpha+1} - \theta_{\alpha} - 2\theta_{\alpha-1} + \theta_{\alpha-2}) + \cos\frac{1}{2}(\theta_{\alpha+2} - \theta_{\alpha+1} + \theta_{\alpha} - 2\theta_{\alpha-1} + \theta_{\alpha-2})$$

$$)$$

(A9) $$\tau = 8\sin\frac{1}{2}(\theta_{\alpha} - \theta_{\alpha-2})\sin\frac{1}{2}(\theta_{\alpha} - \theta_{\alpha-1})\sin\frac{1}{2}(\theta_{\alpha+1} - \theta_{\alpha})\sin\frac{1}{2}(\theta_{\alpha+2} - \theta_{\alpha})$$

(A10) $$g_{\alpha\sigma} = \frac{1}{\tau}\sin\frac{1}{2}(\theta_{\alpha+2} + \theta_{\alpha+2} + \theta_{\alpha-1} + \theta_{\alpha-2})$$

Appendix B: Solution for Optimal 2nd Harmonic

We require the best possible match between the energy angle and the pressure angle. One way of doing this is to minimize the following functional:

(B1) $$\left(\frac{\sum g_i^2 \sin(\theta_i)}{\sum g_i^2 \cos(\theta_i)} - \frac{\sum g_i \sin(\theta_i)}{\sum g_i \cos(\theta_i)}\right)^2$$

At first this may seem intractable. If we note that any solution to equation (5) has the property that the pressure angle is exactly φ, then we may restate the functional as follows:

(B2) $$\left(\frac{\sum g_i^2 \sin(\theta_i)}{\sum g_i^2 \cos(\theta_i)} - \frac{\sin(\varphi)}{\cos(\varphi)}\right)^2$$

With some rearrangement, we can produce another functional that is largely equivalent:

(B3) $$\left(\sum g_i^2 \sin(\theta_i - \varphi)\right)^2$$

Note that none of the three functionals is exactly equivalent to minimizing the difference between the energy angle and the pressure angle. Since in most cases, the match is exact, all three functionals will go to zero at an exact match, so the desired effect will be achieved.

Next, we express the gains, g_i , as a combination of the solution with no 2nd harmonic and the solution with the 2nd harmonic equal to one.

(B4) $$g_i = \alpha_i + \sigma \beta_i$$

where α_i is the speaker gain with zero second harmonic, β_i is the speaker gain with unit second harmonic, and σ is the amount of second harmonic (which is unknown at this point). We substitute (A4) into (A3), collect terms, and we have the following:

(B5) $$\left(\sum (\alpha_i^2 \sin(\theta_i - \varphi)) + 2\sigma \sum (\alpha_i \beta_i \sin(\theta_i - \varphi)) + \sigma^2 \sum (\beta_i^2 \sin(\theta_i - \varphi))\right)^2$$

The term being squared is a 2nd-order polynomial in the unknown σ. Let us denote that polynomial by $p(\sigma)$. We can then express the functional and the condition as follows:

(B6) $$p^2(\sigma) = \min$$

If we apply the usual least-squares technique, we differentiate the above equation to obtain something we can solve for σ:

(B7) $$p(\sigma)p'(\sigma) = 0$$

Any time you have the product of two polynomials equal to zero, one or the other has to be zero (or maybe both). This reduces then to two separate equations:

(B7a) $p(\sigma) = 0$

(B7b) $p'(\sigma) = 0$

This gives us three solutions. If (B7a) yields complex solutions, we ignore them and choose the solution from (B7b). We will then have either 1 or 3 real solutions. We may choose any one of them. In the case that there are 3 real solutions, we may make it unique by requiring, for instance, that the Euclidean norm of the resulting gain vector (equation (B4)) be a minimum.

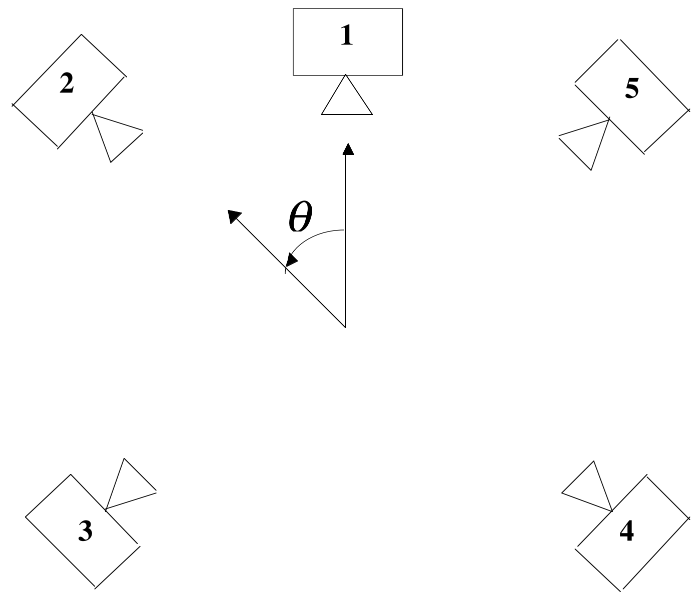

Figure 1: Numbering scheme for speakers and definition of angular position. We assume that the speaker layout is symmetric about the front-back axis.

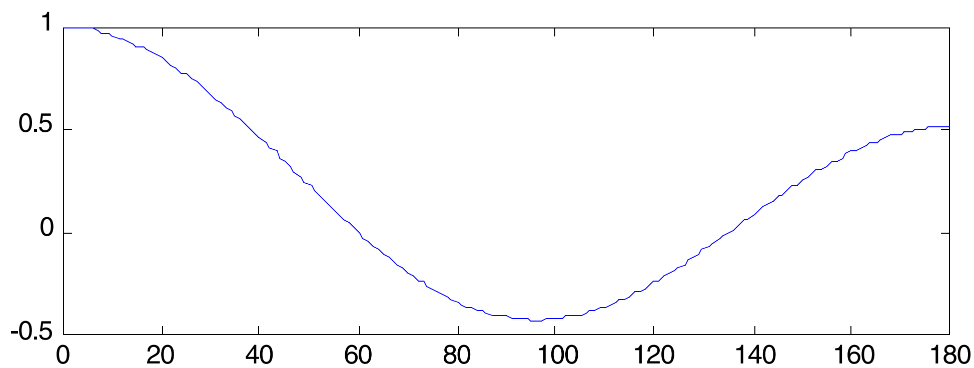

Figure 2: Gain to front speaker using 0th, 1st and 2nd order harmonics. Note the undesirable behavior. The gain is bimodal. As the direction of the sound approaches 180° (behind the listener), the gain rises to a secondary maximum. Generally we want the speaker gains to be smooth and unimodal as the sound position moves around the listener.

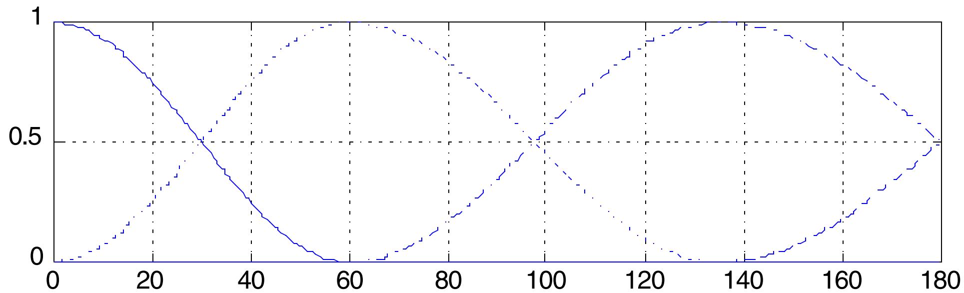

Figure 3: "Standard" pan formula with -6dB crossover points, extended to 5-channel usage.

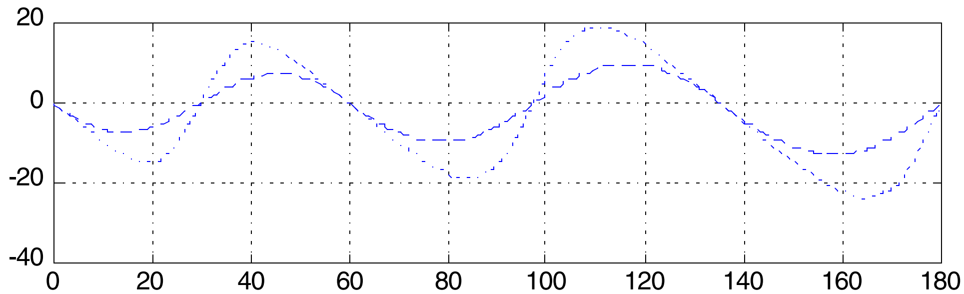

Figure 4: For "standard" gain with -6dB crossover points, deviation of pressure and power angle from true are plotted.

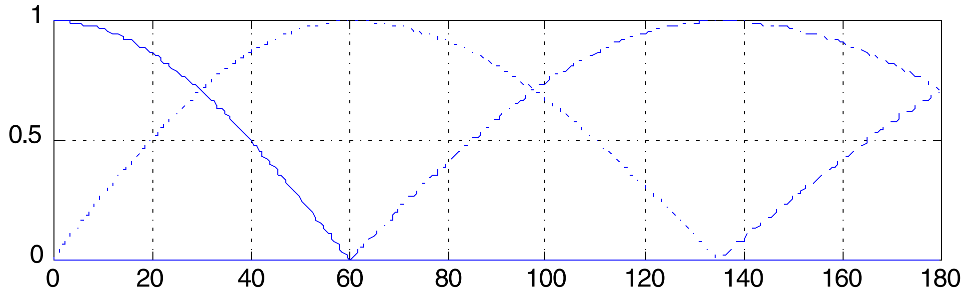

Figure 5: "Standard" pan formula with -3dB crossover points, extended to 5-channel usage.

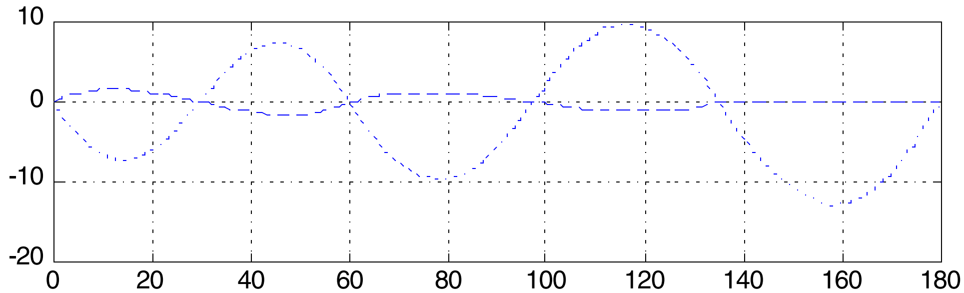

Figure 6: For "standard" gain with -3dB crossover points, deviation of pressure and power angle from true are plotted.

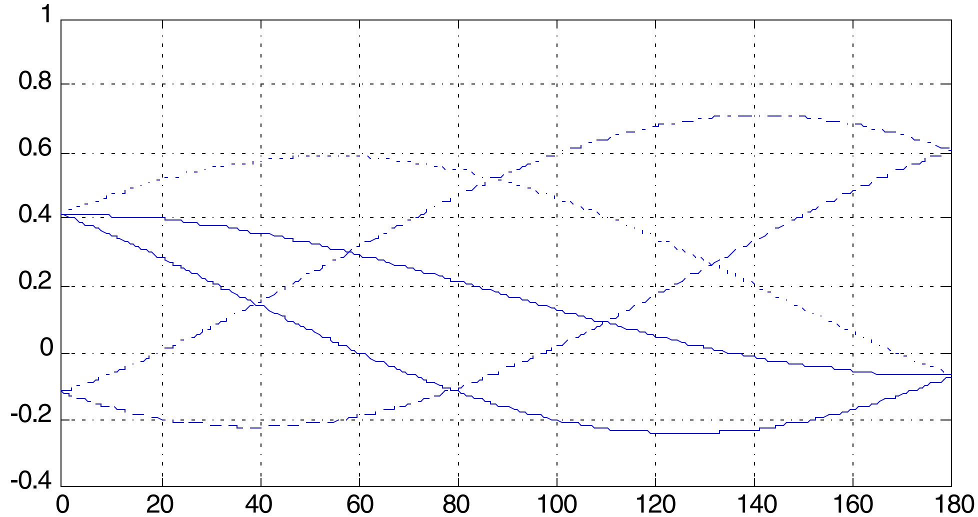

Figure 7: 5-channel gains versus virtual source angle. Gains are constrained to have zero second spatial harmonics

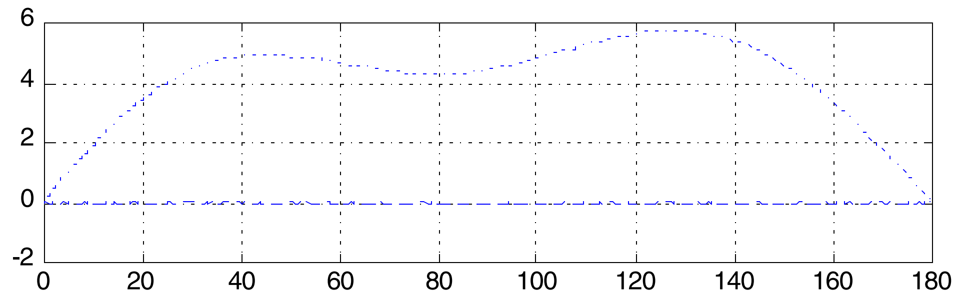

Figure 8: For gains constrained to have zero second spatial harmonics, deviation of pressure and power angle from true are plotted. Of course, the pressure angle is exact, so there is no deviation.

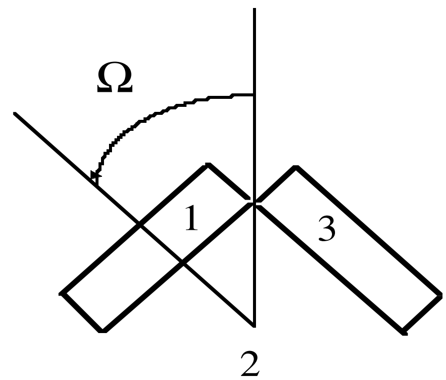

Figure 9: A 2-dimensional sound field microphone consisting of 3 identical directional capsules. They may be coincident, or the forward-facing capsules may be separated as is used in the ORTF arrangement (in which case, the capsules should be hypercardiods).

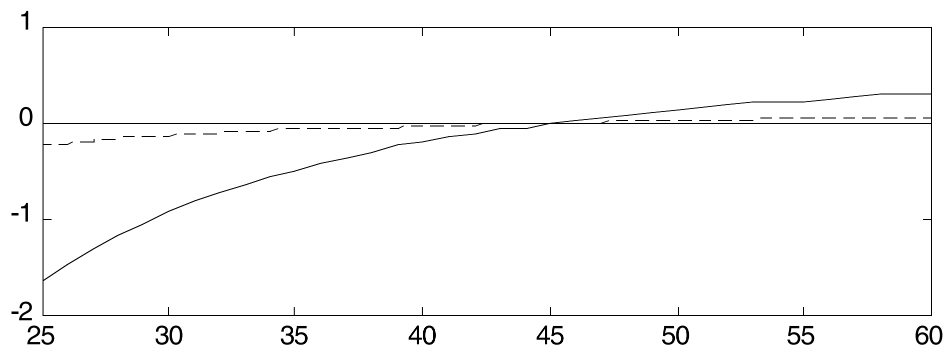

Figure 10: Coefficient of rear-facing microphone (closest to zero) and either forward-facing microphone to the center speaker for varying angles of the front speakers.

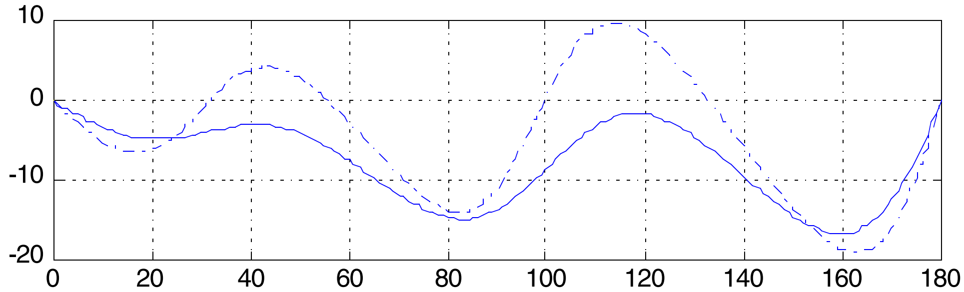

Figure 11: Pressure and power angle (in degrees) deviation from true direction of sound source for pickup consisting of five omni-directional microphones.

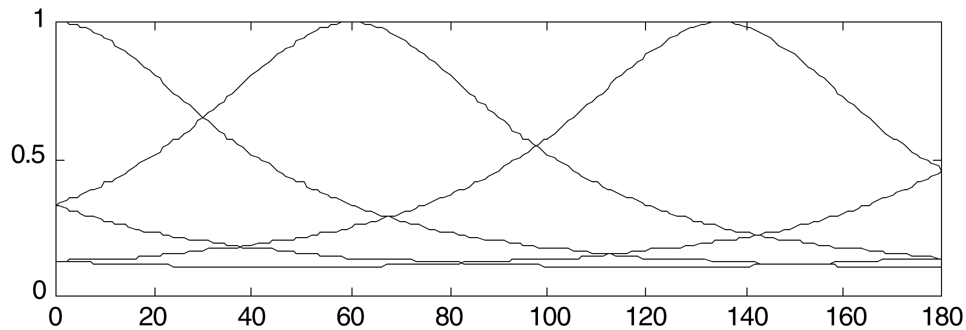

Figure 12: Signal strengths to 5 omni-directional microphones as a sound is moved around them.

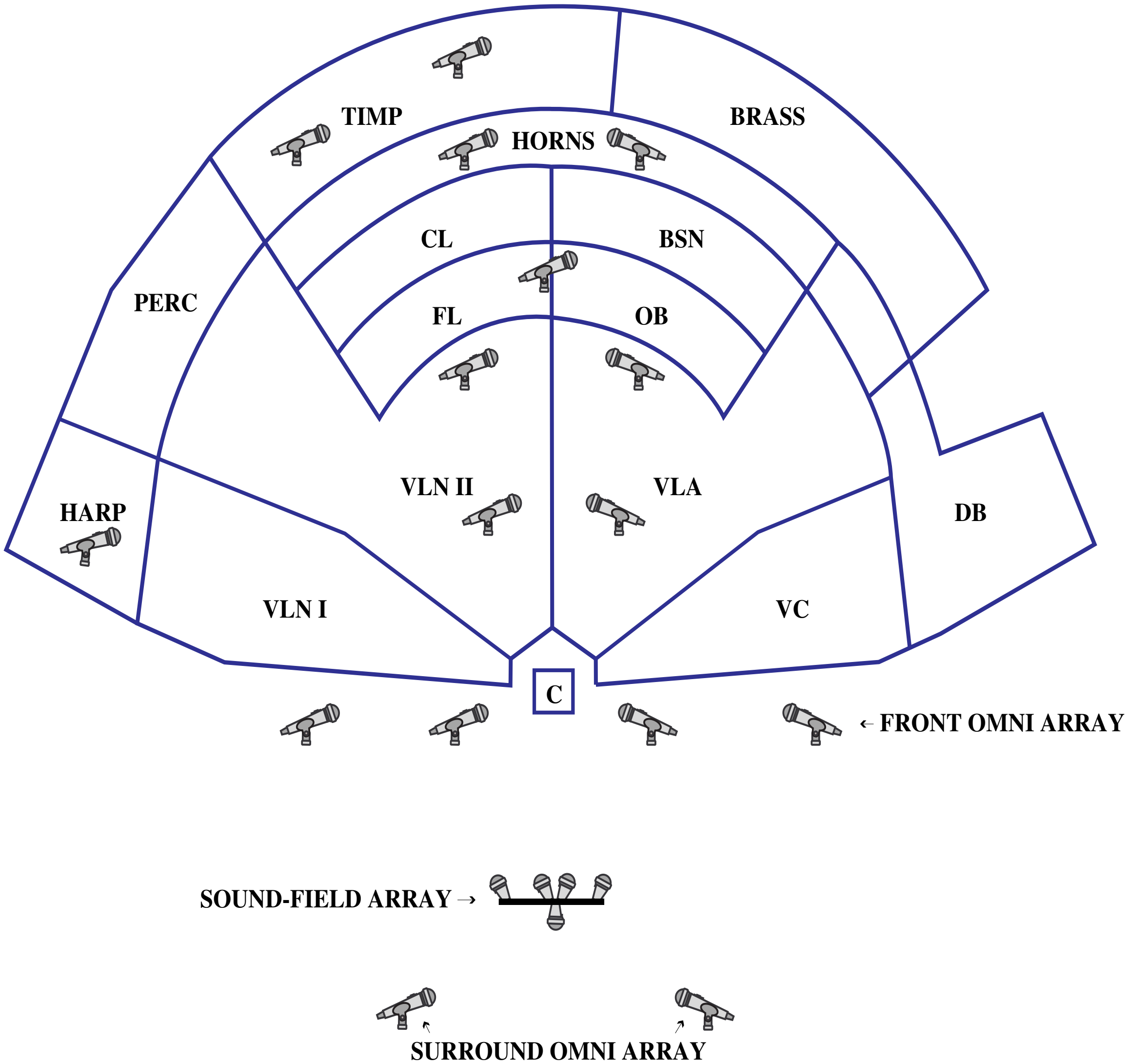

Figure 13: Simplified schematic diagram of the orchestra placement showing the locations of the microphones. The "spot" microphones were a mixture of cardoid and hypercardoid directional patterns. The sound-field array consisted of one X-Y (coincident) pair of hypercardoid microphones at an angle of 109.47° degrees, one ORTF pair, and one rear-facing hypercardoid.