

MUSIC RECORDING IN THE AGE OF MULTI-CHANNEL

James A. Moorer
Sonic Solutions

ABSTRACT:

The DVD-Video 1.0 standard allows a disk that has little or no video on it, but can carry multiple channels of PCM audio. This allows us to provide an enhanced spatial experience by using more channels than the traditional stereo recording. We propose a number of new processes that may be incorporated into the production chain, from microphone design to improved pan matrices to corrections in the home for speaker placement. These new processes start to take advantage of the possibilities of the format. We suggest that the analytic framework of spatial harmonics can provide a rational way to unify and evaluate choices at each stage of the process.

INTRODUCTION:

There have been any number of proposals for speaker and microphone placement over the years, and any number of suggestions for increasing the number of channels of audio that are distributed with commercial recordings. With vinyl records as the means of distribution, methods of delivering more than two channels were dubious, at best. Since the mid-1960s, film sound has routinely used multiple speakers placed around the theater. Since the 1970s, this has become largely *de facto* standardized as the modern LCRSS placement (left, center, right, left-surround, and right-surround).

If one asks the question “how many speakers are really necessary?”, the theoretical answer is somewhat unhelpful: to exactly recreate a wave-front, we need the analogy of an “acoustic hologram”, which would require a 2-dimensional array of speakers located at spacings no greater than $1/2$ the wavelength of highest frequency desired. This is somewhat less than 1 cm. For a reasonable-sized theater, this would require more than 20,000,000 independent channels. It is capable, however, of exactly recreating a particular wavefront up to the range of human hearing. Even for a home theater, it is on the order of 4,000,000 channels to cover one wall. We might want to cover, say, 6 surfaces for “total immersion” in a perfectly recreated wavefront. We will not pursue this line of reasoning any further in this paper.

With the introduction of DVD-Video, we have the chance to reevaluate the question of multi-channel music recording. Pragmatic arguments, such as what the consumer is likely to go for, and the need for stereo CD releases for the foreseeable future should play an important part in our discussion. Similarly, DVD is only the one stage in the evolution of vehicles for the distribution of audio and video content. Any casual observer of technological progress will know that it will not be the last stage.

BASIC ASSUMPTIONS:

Although it might be nice to hope that consumers would place speakers wherever we might wish, in fact the speaker placement will be strongly influenced by the DVD-Video standard, which requires film-type LCRSS placement, as shown in Figure 1. We might like to place speakers above the listener for full periphonic reproduction [1], but that would reduce the number of households that would be able to enjoy such recordings. Consequently, for the purposes of this exposition, we will take a number of postulates to be the starting point and discuss what can be done even with these limitations. One could, of course, start with a different set of postulates. We assume that we have complete control on the production side of the process.

- We take the speaker placement of Figure 1 as a given, with some uncertainty in the exact locations of the speakers.
- We assume that the speaker arrangement is symmetric about the front-back axis.
- We accept that generally there is less audio power available to the rear speakers than to the front speakers. Additionally, there may be some coloration in the rear speakers.
- We assume that the 3 front speakers are identical.

With these premises, we can now discuss what can and can not be accomplished on a rational basis.

We note here that there are a number of differences between the surround systems in motion picture theaters and those in homes. The most obvious one is scale. If we take the speed of

sound to be roughly 1 foot per millisecond, then we note that theater viewers are already 20 milliseconds or more from the nearest speaker. Just moving from one seat to the next may add a millisecond to one path and subtract a millisecond from another. Two different listeners may be seated 20, 30, 40 or more milliseconds apart. In the home, however, the entire listening area will be maybe 10 milliseconds wide. This makes it possible to use low-frequency phase information for additional localization cues, as will be noted later.

THEORY OF DIRECTIONALITY:

We will summarize briefly some aspects of directional theory. We use the definition of angle and speaker number shown in Figure 2. 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. Later on, we will use the same coordinate system for a sound-field microphone. This microphone is also placed at the exact center of the coordinate system. Since both the human head and a physical microphone have non-zero size, we will assume that they are centered somehow on the coordinate system.

Gerzon [2] presented a summary of some aspects of directional theory that included calculation of the Makita (velocity) localization vector and the power¹ 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 [3]. We may define the angle determined by velocity, θ_V , and the angle determined by power, θ_P , as follows:

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

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

$$(2) \quad \theta_P = \tan^{-1} \left(\frac{\sum g_i^2 \sin \theta_i}{\sum g_i^2 \cos \theta_i} \right)$$

It is desirable that these vectors coincide, since that would then present a coherent image where the perception of high and low frequencies correspond. Note that these vectors do not correspond exactly to the way human hearing determines direction, but they at least give us some kind of analytic “handle” on the problem of evaluating systems of spatialization.

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:

$$(3) \quad S_i = W + Y \cos \theta_i + X \sin \theta_i$$

¹ Gerzon referred to this is the “energy” vector, although it is more precisely called a “power” vector.

Again, θ_i is the angle of speaker i using the coordinate system shown in Figure 2. 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°).

It is straightforward to show that with the speaker signals shown in equation (3), using 4 speakers in a perfect square, the velocity and power vectors correspond. Gerzon also generalized the theorem to any regular polygon [4]. 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 (3) will allow the velocity and power 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 Figures 1 and 2, the velocity and power vectors can not be made to align except at a number of discrete values. These include, of course, the directions towards the speakers themselves, but alignment can be achieved at some finite number of points between the speakers. This is the best we can do.

PLACEMENT OF A SOUND:

Since many releases are made using multiple, separately recorded tracks, we may begin by asking the question "How do we distribute the sound among the speakers to give the impression of a sound at a particular location"?

To my knowledge, Gerzon never posed this particular question, although the answers we derive will fall directly out of his formulation for directionality. From this point on, all the results are new, though heavily based on Gerzon's formulation.

To answer the question, we may appeal to the Fourier sine and cosine series on the circle. This is equivalent to the spin harmonic functions described in [1], 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:

$$(4) \quad \frac{1}{2} + \sum_n \cos n(\theta - \phi)$$

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.

This formula is derived in Appendix A.

The contribution to directivity from speaker i will be:

$$(5) \quad g_i \left\{ \frac{1}{2} + \sum_n \cos n(\theta - \theta_i) \right\}$$

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 (4)) to the directivity function obtained by our speaker placement (equation (5)). This fit may be performed as a least-squares operation to determine the unknown channel gains³. After some manipulation, we arrive at the set of linear equations as follows:

$$(6) \quad 1 + 2 \sum_n \cos n(\phi - \theta_k) = \sum_{i=1}^N g_i \left\{ 1 + 2 \sum_n \cos n(\theta_i - \theta_k) \right\}$$

for $k = 1, 2, \dots, N$. This gives N equations in the N unknown channel gains. In our case, of course, $N = 5$.

Note that this result can also be obtained simply by setting equation (4) equal to the sum of equation (5) for each of the speaker angles.

The astute reader will notice that the bounds on the summation on the cosine series have been routinely omitted so far. This is deliberate. Technically, the expansion is not limited. Since it is non-convergent, it makes it 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.

This is a mixed blessing. On the one hand, it says that we cannot hope to achieve a high degree of directionality, even with 5 speakers, since we can only recreate the zero-th and first spatial harmonics. On the other hand, for a 4-speaker rectangular setup, these equations can reduce to the familiar pan-pot relations with the gain exactly 3dB down when the sound is exactly between two speakers. With 4 or more speakers, a signal coming directly from a particular speaker can have zero gain in the other speakers. I find that comforting.

³ Just to make it explicit, subtract (4) from (5), square it, and integrate with respect to θ over 2π , then differentiate with respect to the unknowns and set the result to zero.

There is one difference between the solution for channel gains given by equation (6) and what you might imagine for surround-sound pan pots, and that is when a sound is between two speakers, there are contributions from *all* the speakers. This is somewhat non-intuitive but it is required to preserve the spatial harmonics.

Notice that the right side of equation (6) depends only on the speaker placement. This matrix can be computed and inverted once. 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.

Equation (6) has rank 3 for any number of speakers greater than 3. This is due to the fact that we only have 3 free parameters when we take only the zeroth and first spatial harmonics. To get a solution, other constraints must be applied. For instance, we may add one constraint simply by requiring that the sum of all the speaker gains be unity.

Equation (6), even when augmented by additional constraints, exhibits symmetries that may be exploited to speed up the solution. We will not discuss this further in this paper, but we will refer the interested reader to Golub and Van Loan [5].

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 (6) to determine the channel gains.

There is an interesting additional use for equation (6), 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. All we do is set ϕ in equation (6) to the original speaker positions. This gives us N sets of gains, which relate the original speaker feeds to the new speaker feeds.

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.

The same procedure can be used to rotate the entire sound field. Using panning matrices derived from equation (6), we can smoothly rotate the sound field so that, for instance, what was in the center speaker is now coming from the rear, and so on.

RECORDING FOR MULTI-CHANNEL:

So far, the discussion has concerned itself with studio-produced sound, where the artist will place the sounds in particular locations to achieve the desired aesthetic effect. The missing link is some description of how we can record live performances to take advantage of the multiple-speaker possibilities. Again, carrying forward the notion of spatial harmonics, we will suggest a 2-dimensional sound-field microphone as shown in Figure 3. This uses three identical cardioid capsules arranged in an equilateral triangle. The entire microphone must be as small as possible to reduce high-frequency phasing effects, and the capsules must be matched as well as possible for frequency and phase response (otherwise the matrixing indicated in equation (8) below produces audible artifacts). The individual feeds from the capsules will have the following spatial patterns:

$$(7) \quad \begin{aligned} m_1 &= 1 + \cos\left(\theta - \frac{\pi}{3}\right) \\ m_2 &= 1 + \cos(\theta - \pi) = 1 - \cos(\theta) \\ m_3 &= 1 + \cos\left(\theta + \frac{\pi}{3}\right) \end{aligned}$$

With simple combinations of these, we can produce the coefficients of a first-order spatial harmonic expansion as follows:

$$(8) \quad \begin{aligned} a_0 &= \frac{1}{3}(m_1 + m_2 + m_3) \\ a_1 &= \frac{1}{3}(m_1 - 2m_2 + m_3) = \cos\theta \\ b_1 &= \frac{1}{\sqrt{3}}(m_1 - m_3) = \sin\theta \end{aligned}$$

Since the spatial harmonics are orthogonal, we know that this arrangement of capsules will pick up only the zero-th and first spatial harmonics. All other higher harmonics will cancel out in the matrixing process of equation (8).

We can then use these pickup patterns to solve for the speaker feeds, s_i by an equation similar to equation (6):

$$(9) \quad a_0 + 2a_1 \cos(\theta_k) + 2b_1 \sin(\theta_k) = \sum_{i=1}^N s_i \{1 + 2 \cos(\theta_k - \theta_k)\}$$

Since the right side depends only on the speaker placement, we may easily solve for the required signal feeds. By replacing a_0 , a_1 , and b_1 by the microphone feeds from the two-dimensional sound-field microphone, we have the entire matrix to convert directly from the microphone feeds to the speaker feeds that guarantees the zero-th and first spatial harmonics are preserved.

Note that this again reduces precisely to Gerzon's rectangle decoder theorem in the case of equi-angular speaker placement, as shown in equation (3). Since the matrix implied in (9) is also rank-deficient, we can use additional constraints to get a solution, such as requiring the sum of the speaker feeds to equal the sum of the original microphone feeds, and so on. The rank-deficiency of this matrix shows that there are only 3 independent channels of information, which is what you would expect, given that there are 3 microphone feeds.

Since we also know how to rotate the spatial field, as pointed out in the previous section, we have a number of interesting possibilities. For instance, the sound-field microphone

could be placed directly in the center of a string quartet. The listener, sitting at home, could then rotate the field to position any particular instrument at the front, or into a particular speaker. For an orchestral recording, however, the utility of rotating the sound field so that the audience is at the front is dubious.

There is a somewhat startling corollary to this exposition: *the matrixing described in equation (9) does not need to be applied in the studio at the time the recording is mastered: it can be applied in the home.* The amount of DSP necessary to compute and perform the matrixing is easily available in modern home electronics.

This means that for this kind of recording, it is not necessary to transmit all 5 channels (or stamp them into a DVD). It is only necessary to transmit the individual components a_0 , a_1 , and b_1 , then construct the 5 speaker feeds directly from these, since there is redundancy inherent among the 5 channels.

This does not work for film sound or historical multi-channel material, since the pan matrices were not designed to preserve the spatial harmonics of the signals. For new recordings, we can contemplate recording on a DVD, say, 3 channels of 24-bit 96 kHz PCM, then matrix it into the exact home speaker placement at the time of presentation. The data rate is 6.912 Mbits/sec, which is well within the bandwidth available on the current family of DVD hardware.

Again, this was suggested by Gerzon [2] twenty years ago (in the context of quadraphonic playback)! He also suggested a number of different ways of reducing these three channels down to two for stereophonic compatibility. He pointed out that since there is no optimal way to do this, various different tradeoffs can be employed, depending on what aesthetic result is desired, which then produce a number of different acceptable matrices.

ANOTHER 2-DIMENSIONAL SOUND-FIELD MICROPHONE:

As we know, stereo recording and playback will be with us for some time to come. As noted before, Gerzon outlined a number of ways to reduce the sound-field recordings to stereo which take various explicit compromises in the process. I would like to suggest here a modification to the 2-dimensional sound-field microphone that produces a stereo recording without matrixing simultaneously with a sound-field recording. Figure 4 shows a modified sound-field microphone that is in the shape of an isosceles triangle. Capsules 1 and 3 are at an angle of 90° . Obviously, the feeds from capsules 1 and 3 are exactly following the Blumlein crossed-microphone technique, which is a popular method in use today. Recording the third channel simultaneously allows the matrixing required for surround-sound reproduction that preserves the zeroth and first spatial harmonics. In the stereo case, the rear-facing channel can simply be dropped (or mixed into the other two channels if desired). The matrixing equations are changed slightly as follows:

$$(10) \quad \begin{aligned} m_1 &= 1 + \cos\left(\theta - \frac{\pi}{2}\right) \\ m_2 &= 1 + \cos(\theta - \pi) = 1 - \cos(\theta) \\ m_3 &= 1 + \cos\left(\theta + \frac{\pi}{2}\right) \end{aligned}$$

The coefficients of a first-order spatial harmonic expansion can be calculated as follows:

$$(11) \quad a_0 = \frac{1}{2 + \sqrt{2}}(m_1 + \sqrt{2}m_2 + m_3)$$

$$a_1 = \frac{1}{2 + \sqrt{2}}(m_1 - 2m_2 + m_3) = \cos \theta$$

$$b_1 = \frac{1}{\sqrt{2}}(m_1 - m_3) = \sin \theta$$

These can be used directly in equation (9) to compute the speaker gains for surround sound. Since equation 9 depends only on the angles of the speakers in the listening environment, it can be computed independently of the

Again, if we transmit a_0 , a_1 , and b_1 , we can easily recover m_1 and m_3 by matrixing to produce a stereo feed from the Blumlein crossed-pair capsules, or the three channels can be matrixed to 5 speakers using equation (9) for a surround-sound presentation.

BEYOND THE BEYOND:

The above discussion is oriented towards achieving realism. That is not always the goal. Sometimes, it is desired to go beyond the realistic. This is especially true in major motion-picture production. Often the effects are required to be “bigger than life”. There are a number of ways that this can be done using the advantages of multi-channel presentation.

One effect that can be used in the home, but is not effective in the theater is the concept of driving opposing speakers out of phase. This is only effective for the signal at frequencies below about 500 Hz. It has the effect of forcing the image “beyond” the speakers. By using shelving filters, we can cause low frequencies to be driven out-of-phase, but preserve phase coherence for high frequencies while still expanding the size of the perceived space. As noted above, equation (6) is rank-deficient. Although one may see this is a liability, it is also an asset. As noted above, one might constrain the gains so that they sum to unity. There is still one free parameter that will give a family of valid solutions. This parameter can be used, for instance, to force all the speaker gains to be non-negative. For special effects, however, the free parameter can be used to force some speakers out-of-phase (negative gain). We might solve the gain matrix twice: once for low frequencies, where we force some speakers out-of-phase, and once at high frequencies, constraining all the gains to non-negativity⁴, then using zero-phase shelving filters to make frequency-dependent pan matrices.

Chowning [6] described a system for simulating distance cues by reverberation and motion cues by Doppler shift. This technique is applicable to our speaker placement by simply changing the pan matrices to match those produced by equation (6). For applications of sound with picture, time codes can be assigned to points along the trajectory so that the

⁴ It does not make sense to use negative gains at high frequencies, since the human ear will not be able to make use of this information. The wavelengths are so short that even slight movements of the listener cause the absolute phasing to reverse. For instance, at 10 kHz, a movement of 1/2 inch towards a particular speaker will reverse its phase.

sound will be at a certain position at a particular video frame. Again, if we know that the sounds were positioned by using equation (6), then we have a rational way to redistribute the sound to take care of speakers in different locations from the original mastering studio.

Chowning noted that in human hearing, the ratio of the strengths of the direct signal and the reverberant signal can be used to provide a distance cue. A very clean, dry signal is perceived to be “close” to the ear, whereas one that is full of reverberation is perceived to be “far” from the ear. Chowning had the strength of the reverberant field be roughly constant, then made the direct signal louder or softer to set the distance to the listener according to the $1/r^2$ law.

Chowning did not include in his pan matrices a way to bring the trajectory of a sound *inside* the area delimited by the speakers. Using equation (6), there is no trick to it, since a sound at a certain angle will dictate a series of channel gains regardless of the distance to the virtual sound source. Of course, a literal interpretation of the $1/r^2$ law for the strength of the direct signal at a distance r from the listener would force an infinite gain as the virtual sound source passes through the listener. The gain can be bounded simply by using the (non-realistic) formula $1/(r + \alpha)^2$ instead. This will constrain the maximum gain to be simply $1/\alpha$.

Doppler shift is useful for sound effects, but not for musical sound. If we choose to fly music around the room, we should probably suppress the Doppler shift to preserve correct intonation. For sound effects, we generally want to exaggerate the Doppler shift well beyond what the actual velocities would produce, so controls should be made available to the user to select “realistic” settings, or arbitrarily exaggerated settings.

There are practical limits on the speed with which a sound can be “flown” around the room. If the sound switches from one speaker to the next too quickly, it starts to be perceived as being amplitude-modulated rather than simply placed in space. This is because the pan functions can change so rapidly that they produce sidebands (modulation products) that will become audible.

VIOLATIONS OF BASIC ASSUMPTIONS:

We noted at the beginning of the paper that there were certain assumptions driving the arguments and analysis presented. We should say a few words about the effects of violations of these assumptions:

(1) *The speakers are not all identical.* Although your local hi-fi store would love to sell you 5 identical speakers, generally we can expect the rear speakers (surrounds) to be of different manufacture and different quality. They will have coloration that is different from the front speakers and different power levels, and different distances to the listener. Obviously, we can compensate for distances by simple time delays (or advances) and gain corrections. Frequency-dependent compensation for mismatch between speakers can be contemplated, but we cannot expect this in all households. This is probably the biggest limiting factor to the usefulness of this system. Frequency-dependencies will cause localization of different frequency ranges to differ. The ultimate effect will be to diffuse the image.

(2) *The listener is not always in the exact center.* This is always a problem, even with 2 speakers. The techniques mentioned here have “reasonable” tolerance for listener placement, but there is definitely a “sweet spot.” One could argue that the redundancy

implied in the 5 channel presentation of 3 independent signals gives better coverage, so that the 5 channel presentation will generally have better listener independence than stereo, but there are limits and there will be degradation of the imaging as the listener moves out of the sweet spot. Note again that by adjusting the matrix gains and the time delays to the speakers, the sweet spot can be moved around the room, but it is still limited to the area surrounding a single point in the room.

(3) *The speakers are not at the same height.* The entire derivation above was assuming the speakers and the listeners' ears are all in the same plane. Of course, they are not. In general, the surround speakers will have different heights than the front 3 speakers. Again, this will degrade the imaging. If some speakers are too far out of the plane, they can provide cues for the ear's height perception. This has the effect that the speakers that are out of the plane will no longer fuse with the in-plane speakers and will be heard independently.

Note that in each case, the form of the degradation is less precise imaging. It is not coloration, phasing or distortion (except in the case that we are using frequency-dependent gains or out-of-phase drive - then both coloration and phasing can occur). One could argue, in fact, that the coloration and phasing inherent in having more than 1 sound source (due to cancellation and reinforcement of wavefronts as the listener moves around the room) is probably better with 5 speakers using just about any kind of pan matrices than it is in stereo. In stereo, there are positions where tones from two speakers will exactly cancel or exactly reinforce. With non-zero gains in 3 or more speakers, there are only a few positions where the tones will cancel exactly. The more non-zero gains there are, the fewer places will have exact cancellation, and the maximum and minimum levels of the signals will converge. To make this more explicit, the use of 3 or more non-zero gains in 3 or more speakers will have the effect of widening the "sweet spot."

CONCLUSIONS:

This discussion shows that there are distinct advantages to using more than 2 channels for music recording and distribution, since higher-order harmonics of spatialization can be recorded and presented. We have shown that there are rational, analytic ways of choosing channel gains to place sounds in space that preserve the zeroth and first spatial harmonics. We can make recordings using a 2-dimensional sound field microphone that captures the sound field in a manner that also preserves the zeroth and first spatial harmonics. There are ways of altering the pan matrix in the electronics in the home to take care of speaker placements that differ from those in the mastering studio. There are also ways to exaggerate the spatial effects for the purposes of producing special effects that are "bigger than life."

All of the above, taken together, show that to make use of the power of multi-channel mass distribution of music that DVD allows, it is necessary to augment the *entire* process of music production, from the recording, to the mixing, to the mastering, and ultimately to the electronics in the home. The techniques described in this paper give a rational and consistent way to do this. There may be other ways of performing these functions, but it is hard to imagine a coherent, analytic basis that is as flexible, in that it dictates everything from microphone design to pan matrix calculation to gain corrections in the home to adjust for speaker or listener positioning. We have shown that a single, mathematical principle (spatial harmonics) can provide exactly this coherent, analytic basis. This is presumably what Michael Gerzon had in mind when he chose the title "The Rational Systematic Design of Surround-Sound Recording and Reproduction Systems".

References:

- [1] Gerzon, Michael A. "Periphony: With-Height Sound Reproduction" J. Audio Eng. Soc., Vol. 21, No. 1, Jan/Feb 1973, pp2-10
- [2] 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).
- [3] 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
- [4] Gerzon, Michael A. "The Rational Systematic Design of Surround-Sound Recording and Reproduction Systems" unpublished manuscript, circa 1975.
- [5] Golub, Gene H., and Van Loan, Charles F. "Matrix Computations" Johns-Hopkins University Press, Baltimore, MD, 1983
- [6] Chowning, John M. "The Simulation of Moving Sound Sources," J. Audio Eng. Soc., Vol. 19, No. 1, Jan. 1971, pp2-5

Appendix A:

The Fourier sine/cosine series represents functions defined on a circle. In our case, we use this series to represent the sound pressure wave incident on a point at any angle θ .

$$f(\theta) = \frac{a_0}{2} + \sum_{n=1} \{a_n \cos(n\theta) + b_n \sin(n\theta)\}$$

For a known function, $f(\theta)$, we can determine the coefficients of this expansion as follows:

$$a_n = \int_0^{2\pi} f(\theta) \cos(n\theta) d\theta$$

$$b_n = \int_0^{2\pi} f(\theta) \sin(n\theta) d\theta$$

The particular function we are interested in is an impulse at a given angle, ϕ .

$$f(\theta) = \delta(\theta - \phi)$$

where $\delta()$ is the Dirac delta function. This represents a single sound source at a particular angle. This yields the following coefficients:

$$a_n = \cos n\phi$$

$$b_n = \sin n\phi$$

If we then substitute this back into the series for $f(\theta)$, we obtain the following;

$$f(\theta) = \frac{1}{2} + \sum_{n=1} \{\cos(n\phi) \cos(n\theta) + \sin(n\phi) \sin(n\theta)\}$$

We recognize the argument of the summation to be the cosine sum-of-angles formula, so we may write this series in the simplified form shown in equation (1):

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

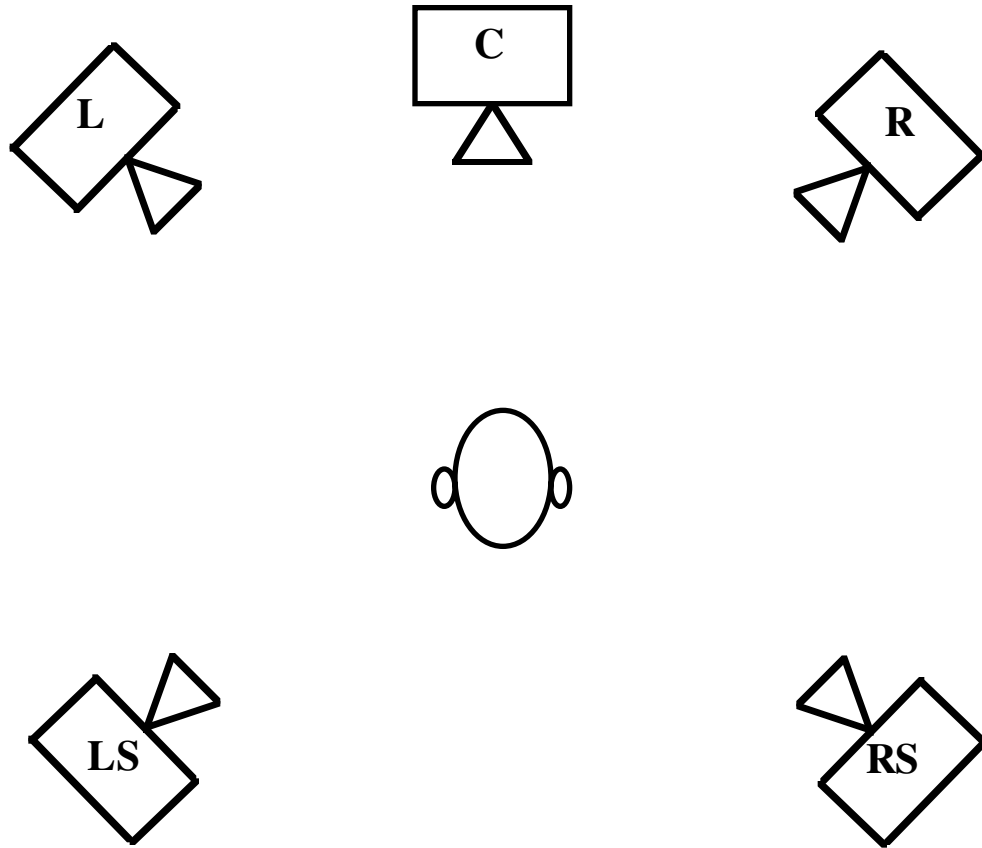


Figure 1: Typical surround-sound speaker placement for home theater applications. The availability of major motion-picture releases in DVD-video format will drive the consumer speaker placement towards this particular layout. Our task, then, is to determine how best to record and master music to take advantage of this particular speaker placement.

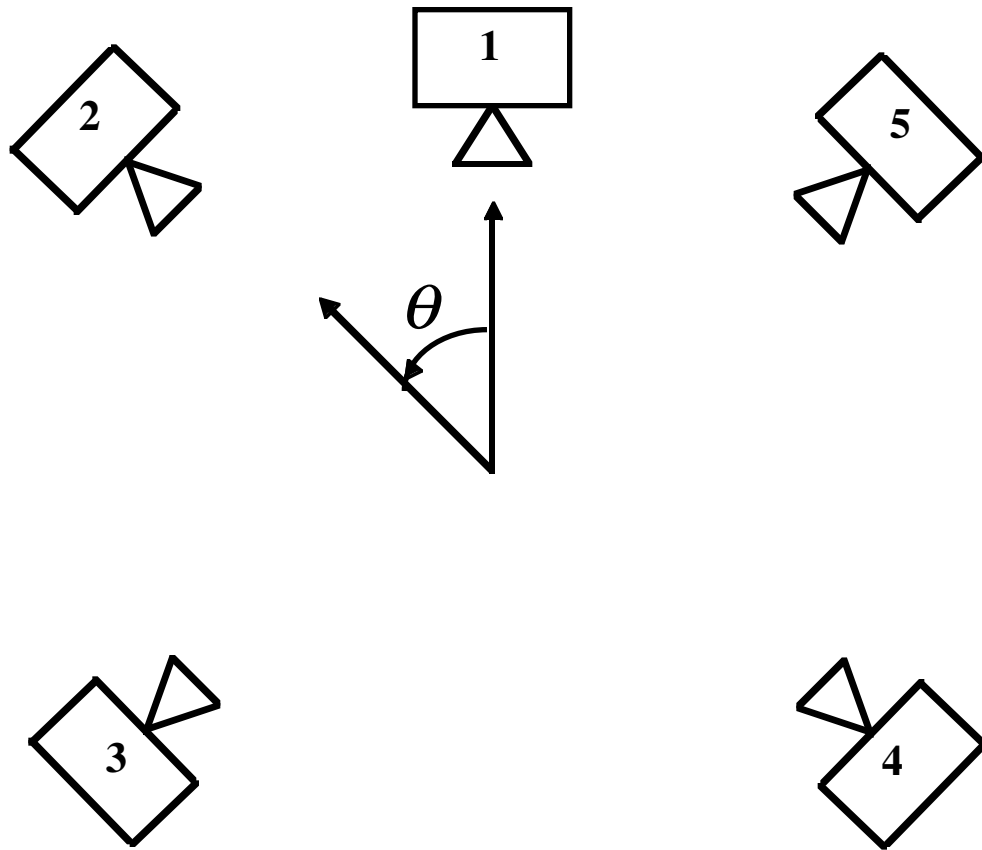


Figure 2: Numbering scheme for speakers and definition of angular position. We assume that the speaker layout is symmetric about the front-back axis.

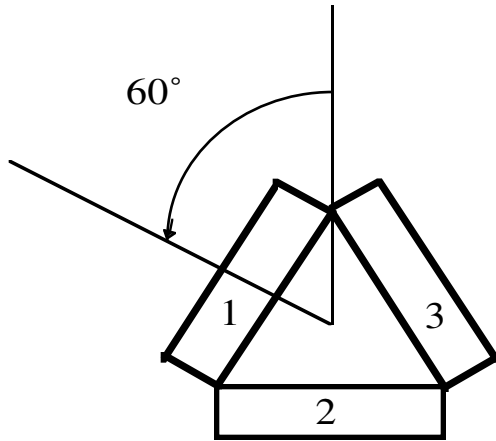


Figure 3: The two-dimensional sound-field microphone. The rectangles represent cardioid-pattern microphone capsules, arranged in a plane with the axes separated by 120° . This is a straightforward special case of the periphonic sound-field microphone, which uses four capsules arranged in a regular tetrahedron. Sums and differences of the feeds from these capsules can give directly the zero-th and first spatial harmonics of the sound field. To limit high-frequency phasing effects, the entire microphone should be as small as possible.

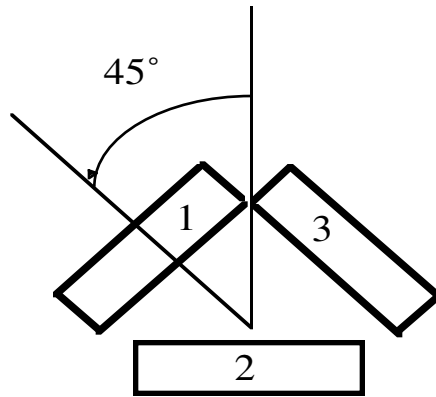


Figure 4: A two-dimensional sound-field microphone that serves two purposes. The rectangles represent cardioid microphone capsules. The three-channel pickup can be used for surround-sound such that the zeroth and first spatial harmonics are preserved. Capsules 1 and 3 can be used separately as a stereo pickup, following the popular Blumlein crossed-mikes technique.