EXTENSION OF THE METHOD OF SPATIAL HARMONICS TO THREE DIMENSIONS

James A. Moorer

We start with the method of spatial harmonics in 2 dimensions (in a plane). The results of this methodology are (1) a way of recording surround sound that can be used to feed any number of speakers (2) a way of panning monaural sounds so as to produce exactly a given set of spatial harmonics, and (3) a way of storing or transmitting surround sound in 3 channels such that 2 of the channels are a standard stereo mix, and by use of the third channel, the surround feed may be recreated that preserves the original spatial harmonics.

This same theory may be extended to 3 dimensions. It then requires 4 channels to transmit the 0th and 1st terms of the 3-dimensional spatial harmonic expansion. It has the same properties for matrixing, such that 2 channels may carry a standard stereo mix, and the other two channels may be used to create feeds for any number of speakers around the listener. Unfortunately, the mathematics for the 3D version is not as clean and compact as for 2D. There is not any particularly good way to reduce the complexity.

To extend the method of spatial harmonics to 3 dimensions, we need to briefly discuss the Legendre functions and the spherical harmonics. In some sense, this is a generalization of the Fourier sine and cosine series. The Fourier series is a function of one angle, . The series is periodic. It can be thought of as a representation of functions on a circle. Spherical harmonics are defined on the surface of a circle, and are functions of two angles, and . is azimuth. Zero degrees is straight ahead. 90 is to the left. 180 is directly behind. is declination (up and down). Zero degrees is directly overhead. 90 is the horizontal plane, and 180 is straight down. Note that the range of is zero to 180 , whereas the range of is zero to 360 (or -180 to 180).

The common definition of spherical harmonics starts with the Legendre polynomials, which are defined as follows:

$$P_n(m) \equiv \frac{1}{2^n n!}\frac{d^n}{dm^n}(m^2-1)^n$$

From these, we can define Legendre's associated functions, which are define as follows:

$$P_n^m(m) \equiv (-1)^m (1-m^2)^{m/2}\frac{d^m P_n(m)}{dm^m}$$

Both the Legendre polynomials and the associated functions are orthogonal (but not orthonormal). We have to specifically give these definitions, since some authors define them slightly differently.

Although these are polynomials, they are turned into periodic functions with the following substitution:

$$m \equiv \cos q$$

From these, an expansion of a function in polar coordinates can be made as follows:

$$f(q,f) = \sum_{n=0} \left\{ A_n P_n(\cos q) + \sum_{m=1}^{n} \left(A_{nm}\cos mf + B_{nm}\sin mf\right) P_n^m(\cos q) \right\}$$

The functions $P_n \cos q$, $\cos mf \; P_n^m(\cos q)$, and $\sin mf \; P_n^m(\cos q)$ are called *spherical harmonics*. This expansion has an equivalence to the Fourier series, but it is relatively messy to actually derive it. What you do is to fix the value of at, say, 90 . The remaining terms collapse into something that is equivalent to the Fourier sine and cosine series.

For a function that is just defined on the circle, there are $1+2T$ coefficients for a series that include harmonics 0 through T. For the spherical harmonic expansion, the total number of coefficients is T^2 .

When applied to sound, this can be though of as the sound pressure on the surface of a microscopic sphere at a point in space. This sphere can be taken to be the location of a listener. We will use this expansion to guide us through the generation of pan matrices and microphone processing for sounds that may originate in any direction around the listener.

We start by taking the function on the sphere that we want to approximate to be a unit impulse in the direction (q_0, f_0) to the listener.

For compactness, we define m_0 as follows:

$$m_0 \equiv \cos q_0$$

We can then state the expansion of a unit impulse in that direction can be calculated to be the following:

$$f_0(q, f) = \sum_{n=0} \frac{2n+1}{2p} \left\{ \frac{1}{2} P_n(m_0) P_n(m) + \sum_{m=1}^{n} \frac{(n-m)!}{(n+m)!} \cos m(f - f_0) P_n^m(m_0) P_n^m(m) \right\}$$

Now let us place N speakers around the listener at angles of $(q_1, f_1), (q_2, f_2), \cdots, (q_N, f_N)$. We then seek gains to each of the speakers, g_i , so that the resulting sound field around a point at the center corresponds to the desired sound field ($f_0(q, f)$ above) as well as possible. We may obtain the gains by requiring the integrated square difference between the resulting sound field and the desired sound field be as small as possible. The result of this optimisation is the following matrix equation:

$$BG = S$$

where G is a column vector of the speaker gains:

$$G = [g_1 \cdots g_N]^T$$

The other terms may be computed as follows:

$$b_{ij} = \sum_{n=0} \frac{2n+1}{2p} \left\{ \frac{1}{2} P_n(m_i) P_n(m_j) + \sum_{m=1}^{n} (-1)^m \frac{(n-m)!}{(n+m)!} \cos m(f_i - f_j) P_n^m(m_i) P_n^m(m_j) \right\}$$

$$S = [b_{10} \cdots b_{N0}]^T$$

Note that this is similar to the expansion of the unit impulse in a certain direction but for the term $(-1)^m$. Although the first summation is written without an upper limit, in practice it will be a finite summation. The rank of the matrix *B* depends on how many terms of the expansion are retained. If the 0^{th} and 1^{st} terms are retained, the rank of *B* will be 4. If one more term is taken, the rank will be 9. Needless to say, the rank of *B* also determines the minimum number of speakers required to match that many terms of the expansion.

Any number of speakers may be used, but the system of equations will be under-determined if the number of speakers is not a perfect square number. There are various ways to solve the under-determined system. One way is to solve the system using the pseudo-inverse of the matrix *B*. This is equivalent to choosing the minimum-norm solution, and provides a perfectly acceptable solution. Another way is to augment the system with equations that force some number of higher harmonics to zero. This involves taking the minimum number of rows of *B* that preserves it rank, then adding rows of the following form:

$$\left[P_{n+1}(m_1)\cdots P_{n+1}(m_N)\right]=[0]$$

or

$$\left[\cos f_1 P^m{}_{n+1}(m_1)\cdots\cos f_N P^m{}_{n+1}(m_N)\right]=[0]$$

or

$$\left[\sin f_1 P^m{}_{n+1}(m_1)\cdots\sin f_N P^m{}_{n+1}(m_N)\right]=[0]$$

It doesn't make much difference exactly which of these are taken. Each additional row will augment the rank of the matrix until full rank is reached.

Thus we have derived the matrix equation required to produce speaker gains for panning a single (monophonic) sound source into multiple speakers that will preserve exactly some number of spatial harmonics in 3 dimensions.

Sound-Field Microphones:

A standard directional microphone has a pickup pattern that can be expressed as the 0^{th} and 1^{st} spatial spherical harmonics. The equation for the pattern of a standard pressure-gradient microphone is the following:

$$f(q,f)=C+(1-C)\left\{\cos\Theta\cos q+\sin\Theta\sin q\cos(f-\Phi)\right\}$$

and are the angles in spherical coordinates of the principal axis of the microphone. That is, they are the direction the microphone is "pointing." The constant *C* is called the "directionality" of the microphone and is determined by the type of microphone. *C* is one for an omni-directional microphone and is zero for a "figure-eight" microphone. Intermediate values yield standard pickup patterns such as cardioid (1/2), hyper-cardioid (1/4), super-cardioid (3/8), and sub-cardioid (3/4). With four microphones, we may recover the 0^{th} and 1^{st} spatial harmonics of the 3D sound field as follows:

$$\begin{vmatrix}A_0\\A_1\\A_{11}\\B_{11}\end{vmatrix}=D\begin{vmatrix}m_1\\m_2\\m_3\\m_4\end{vmatrix}$$

The spatial harmonic coefficients on the left side of the equations are sometimes called *W*, *Y*, *Z* and *X* in commercial sound-field microphones. Representation of the 3-dimensional sound field by these four coefficients is sometimes referred to as "B-format." (the nomenclature is just to distinguish it from the direct microphone feeds, which are sometimes called "A-format").

The terms $m_1, \cdots, m_M$ refer to M pressure-gradient microphones with principal axes at the angles $(\Theta_1, \Phi_1), \cdots, (\Theta_M, \Phi_M)$. The Matrix D may be defined by its inverse as follows:

$$D^{-1} = \begin{vmatrix} C_1 & (1-C_1)\cos\Theta_1 & (1-C_1)\sin\Theta_1\cos\Phi_1 & (1-C_1)\sin\Theta_1\sin\Phi_1 \\ C_2 & (1-C_2)\cos\Theta_2 & (1-C_2)\sin\Theta_2\cos\Phi_2 & (1-C_2)\sin\Theta_2\sin\Phi_2 \\ C_3 & (1-C_3)\cos\Theta_3 & (1-C_3)\sin\Theta_3\cos\Phi_3 & (1-C_3)\sin\Theta_3\sin\Phi_3 \\ C_4 & (1-C_4)\cos\Theta_4 & (1-C_4)\sin\Theta_4\cos\Phi_4 & (1-C_4)\sin\Theta_4\sin\Phi_4 \end{vmatrix}$$

Each row of this matrix is just the directional pattern of one of the microphones. There must be exactly four microphones to unambiguously determine all the coefficients for the 0^{th} and 1^{st} order terms of the spherical harmonic expansion. Of course, the angles of the microphones must be distinct (there should not be two microphones pointing in the same direction) and we cannot have all four microphone axes in a plane (since that would provide information only in one angular dimension and not two). In these cases, the matrix is well-conditioned and has an inverse.

Virtual Microphones:

Given this formulation, we may then transform these microphone feeds into another set of "virtual" microphone feeds as follows:

$$\begin{vmatrix} \tilde{m}_1 \\ \tilde{m}_2 \\ \tilde{m}_3 \\ \tilde{m}_4 \end{vmatrix} = \tilde{D}^{-1} \begin{vmatrix} A_0 \\ A_1 \\ A_{11} \\ B_{11} \end{vmatrix} = \tilde{D}^{-1} D \begin{vmatrix} m_1 \\ m_2 \\ m_3 \\ m_4 \end{vmatrix}$$

The matrix $\tilde{D}$ represents the directionality and angles of the "virtual" microphones. The result of this will be the sound that would have been recorded if the virtual microphones had been present at the recording instead of the ones that were used. This allows us to make recordings using a "generic" sound-field microphone, then later matrix them into any set of microphones. For instance, we might pick just the first two virtual microphones, $\tilde{m}_1$ and $\tilde{m}_2$, and use them as a stereo pair for a standard CD recording.

Any non-degenerate transformation of these four microphone feeds can be used to create any other set of microphone feeds, or can be used to generate speaker feeds for any number of speakers (greater than 4) that can recreate exactly the 0^{th} and 1^{st} spatial harmonics of the original sound field.

To matrix the microphone feeds into a number of speakers, we reformulate the right-hand side of the matrix equation for panning as follows:

$$BG = R = R_1 D \begin{vmatrix} m_1 \\ m_2 \\ m_3 \\ m_4 \end{vmatrix}$$

$$R_1 = \begin{vmatrix} P_0(m_1) & P_1(m_1) & -\cos f_1 P_1^1(m_1) & -\sin f_1 P_1^1(m_1) \\ & \cdots & & \\ P_0(m_N) & P_1(m_N) & -\cos f_N P_1^1(m_N) & -\sin f_N P_1^1(m_N) \end{vmatrix}$$

The matrix, R_1, is simply the 0th and 1st order spherical harmonics evaluated at the speaker positions. One must be careful to include the term $(-1)^m$, since that is a direct result of the least-squares optimization required to derive these equations.

<u>Gradient Vector Calculation:</u>

Note also that the pressure gradient vector may also be generalized to 3 dimensions as follows:

$$a = \sum g_i^g \cos q_i$$

$$b = \sum g_i^g \sin q_i \cos f_i$$

$$c = \sum g_i^g \sin q_i \sin f_i$$

Then the angles may be computed as follows:

$$q_g = \tan^{-1} \frac{\sqrt{b^2 + c^2}}{a}$$

$$f_g = \tan^{-1} \frac{c}{b}$$

As usual in spherical coordinates, care must be taken at $q = 0$ and $q = p$, since f becomes indeterminate due to a singularity in the coordinate system.