Part II

AMBISONICS COMES OF AGE

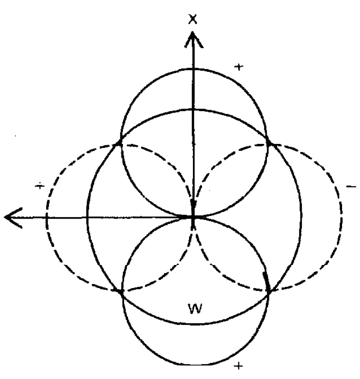
BY WILLIAM SOMMERWERCK

In TAA *3/84, Mr. Sommerwerck provided some background on Ambisonics. He will continue that discussion this time and give more details about the Ambisonic approach.*

MICHAEL GERZON IS a mathematician at Oxford. In the early 1970s, with all the talk of quadraphonics going around, he realized that something was amiss. He saw the need for additional channels to convey the spatial character of live sound more accurately, but he also saw that the pair-wise mixing (PWM) techniques being promoted were incorrect. He was also disappointed with the poor performance of simple matrix decoders and thought that much more credible effects must be possible for listeners who had access only to two-channel program sources. With this in mind, he started theoretical and empirical research into the fundamental problems involved in reproducing directionality correctly. He was joined in this work by Peter Fellgett of the University of Reading. Their research was partially supported by the National Research Development Corporation and IMF Electronics.

Over a ten-year period, Gerzon and Fellgett produced a system that allows recording engineers to achieve almost any directional effect, with total confidence that the listener will hear it correctly. Effects can range from completely natural reproduction of live sound and its ambience to totally artificial studio effects, and these may be freely combined. The system is called Ambisonics.

Its starting point is the same as that for UMX. The sound field is analyzed into four components, one of which is an omni signal without directional information. Three figure-8 components represent front/back, left/right and up/down directionality. These are called the W, X, Y and Z signals, respectively. See *Fig. 7*. (Note that the fourth Ambisonic channel is *not* the "sin 2 theta" used in UMX. Gerzon showed that this signal actually degrades performance. Its presence gives the brain additional cues, which cause sounds *near* a given speaker to localize *at* that speaker. This is undesirable, since the goal is a smooth, continuous spread of sounds, without any loudspeaker awareness. When exaggerated directionality is desired for dramatic emphasis, you can use the fourth channel for just that effect.)



Reprinted with the permission of NRDC.

FIGURE 7: B-format signal arrangement. In this case, the Z signal points up.

The W, X, Y and Z signals can be assembled from live sounds by means of special microphone arrays or in the studio by simply adjusting the amplitude and polarity of the signals, as described earlier. Since the up/down channel is optional, only three channels are needed for normal surround sound.

Besides the formation of a viable surround-sound system, Gerzon and Fellgett studied and applied existing mathematical models of directional hearing. These show that Ambisonic reproduction is correct, and they permit rational trade-offs among conflicting requirements in designing practical systems. Naturally, the researchers performed hundreds of hours of listening tests, but these were guided by mathematics and psychoacoustics. Little is "cut-and-try" in Ambisonic technology.

I have already discussed Makita localization, which is the most important directional cue. The success of any surround-sound system hangs on its proper implementation. Any system that does not satisfy Makita is inherently flawed and cannot accurately reproduce ambience or arbitrary effects. But other psychoacoustic laws must be obeyed, too.

One of these is described by the sound field's "velocity vector magnitude," which determines how stable the field is with respect to head rotation for frequencies below about 1kHz. Changes in this parameter can make the sound move toward, or even into, the listener's head. (Because PWM fails to satisfy this rule, it cannot create stable side images.)

Another rule is described by the "energy vector azimuth," which determines the apparent image location at high frequencies (about 0.5 to 5kHz). (Oddly, PWM *does* work in this range.) Just as the velocity vector magnitude defines image stability for Makita localization, a fourth parameter, called "energy vector magnitude," defines image stability for the energy vector azimuth.

And then there is "phasiness." One example of phasiness occurs when your speakers are connected out of phase. There is no clear center image, and other sounds are bloated and hard to localize. Another symptom is an uncomfortable "pressurein-the-ears" effect, which might range from a slight blurring of image location to the sense that fingers are pushing against your eardrums. Phasiness is related to the amount of 90-degree (quadrature) component in the signal. Only small amounts are acceptable, and the effect appears most pronounced in the 300 to 1,000Hz range.

These effects are described in a group of equations that you can use to analyze any sound-reproduction system, regardless of the number of transmission channels or speakers. I decided that I could not be an Ambisonic "maven" without going through the "maths," so I applied them to the three-channel (W, X, Y) version. They all came out on the button. (The equations, along with an explanation of how to apply them, appear in the sidebar on page 41.) When you analyze the equations, you will see that they account for amplitude and phase differences, but they do not mention arrival time differences. These differences are a significant source of directional information for the ear and brain, especially for the initial transient of any waveform. Why have they been left out of Ambisonics?

You Can't Have Everything

The two reasons for this are simplicity and compatibility. The ear and brain use at least three mechanisms to localize sound—amplitude, phase and arrival time differences. Although the arrival differences *do* make a major contribution, if amplitude and phase are well implemented, the need for delay is less important.

Adding those delays introduces serious problems. You must either use a *kunstkopf* (dummy head—literally, "art head") for live recordings or introduce actual delay for studio recording, via (expensive) delay lines. If you do not want to use headphones in playback, you must supply a crosstalk canceller for loudspeaker listening. This limits the useful "window" to a space suitable for only two people—if they are Siamese twins! When you mix down to mono for broadcasting, the delayed components will "comb" with the undelayed to create frequency-response aberrations. (This might not be significant to those hearing the program over cheap mono players or receivers, but it *is* a potential problem.) The worst part is that you must either wear headphones or install a crosstalk canceller. You cannot sit down and listen to your system in its conventional state.

JVC has combined all the principles of directional hearing, including interaural delay, into a surround system called Q-Biphonics. It works beautifully, but its complexity and incompatibility with existing systems and listening habits will, I believe, prevent it from ever becoming widely accepted. That leaves us with Ambisonics.

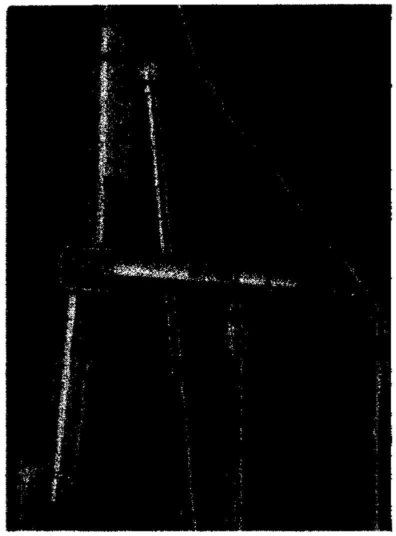


FIGURE 8: The author's home-brewed Ambisonic recording setup, using variable-pattern SoundField microphones.

The Ambisonic Ideal

After all this build-up, just how good is Ambisonics? The three-channel version (horizontal only, no height) is amazing. The sound field has a remarkable coherence that you rarely, if ever, hear in conventional quadraphonic recordings. The ambience and the direct sounds hang together in a way that is difficult to describe, except to say that it sounds almost exactly like the hall in which the recording was made. This is in contrast to quadraphonics, where the ambience is enhanced, but the orchestra sounds too far away, and the reverberation is overdone. Perhaps an example will clarify this point.

I did some of my earliest quad recordings with the Orchestra Society of Philadelphia, an "amateur" group. (A lot of "professional" orchestras should play this well.) I set up four cardioid mikes in prescribed "purist" fashion, arranging them as nearly coincident as possible, with two of them facing the rear. But the recordings still sounded like conventional stereo. Although the mikes were 6 feet behind the conductor, the recording sounded as though I was sitting in row M. And although there was a pleasant increase in ambience, it just did not relate to the direct sounds. If I raised the rear channels to the point where they made a noticeable contribution, they drew attention to themselves. If I reduced their level, the ambience disappeared.

Something was wrong, and now it is obvious what it was. As I explained before, all the side sounds (the significant ambience) were being directionally encoded by amplitude panning between the front and rear mikes—acoustical PWM, if you like. No wonder the ambience was displayed incorrectly in playback!

Legal Home-Brew

After getting the Integrex decoder, I tried a home-brewed SoundField setup *(Fig. 8)*. (The SoundField mike is manufactured by Calrec and distributed by Audio + Design, Calrec, E4480 Hwy. 302, Belfair, WA 98528.) Fortunately, my mikes are variable-pattern, so all I had to do was rig up the necessary stand/boom arrangement to bring an omni and two crossed figure-8s into close proximity.

The recordings were a revelation. Now it seemed as though I was about 10 feet behind the conductor, the reverberation was no long excessive (in fact, the playback was a bit drier than the hall), and everything hung together beautifully. Furthermore, these qualities were audible over a wide listening area. Moving close to one speaker did *not* cause the sound field to collapse in that direction. And although I could not hear specific sounds from the sides, they no longer seemed empty. The sound of the hall at Drexel University filled the listening room, without drawing any attention to itself, just as it would in real life.

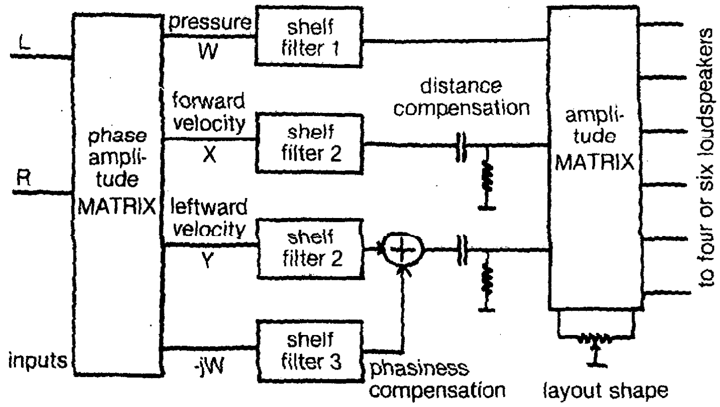
Perhaps the most remarkable and unexpected improvement was the significant reduction in coloration. It was similar to going from Bextrene to polypropylene drivers. A layer of excessively liquid sweetness was stripped away. Woodwinds were noticeably "reedier," violins started sounding more like rosin on gut, and brass instruments had more snap and attack.

This effect seems related to the more correct presentation of ambience that Ambisonics provides. As I have said, the most important ambience is the side delays. In regular stereo recording, these are folded into the direct sound. The brain is incapable of separating them under these conditions, and they "comb" with the direct sound to alter frequency response (and, therefore, instrumental timbre). In Ambisonics, the directional cues necessary for proper localization of the side delays are present, and the combing does not occur.

I made another interesting recording at a neighborhood church's choir rehearsal. The choir was divided, the sections facing each other across the loft. The organ pipes were behind one group, so it was quite natural to have them and one choir to the front, with the other choir "behind" the mike array. The result was marvelous. I could close my eyes and hear the acoustics of the church. (Again, the recording was a bit drier than the real thing.) The speakers disappeared: the recording produced no sense of four sound sources, just a coherent, continuous sound field, no matter which way I turned. It was especially interesting to turn around and pick out individual voices in the rear choir.

Oddly, the organ appeared to be above the front choir, as it was in the church, but I was not using a fourth channel for height, so I chalked this up to autosuggestion. (A British recording engineer told me that although height information is not explicitly encoded, it *might* be there implicitly. The brain might judge the height of a sound source by noting the arrival time difference between direct sounds and those bouncing off the floor or stage. Perhaps this is so, but I do not see how the brain can distinguish between a reflection from the floor and one from the side wall.)

Of course, it will be some time before three and four-channel Ambisonic recordings become available. For better or worse, they will have to await true digital recording techniques. In the meantime, Fellgett, Gerzon and others have developed a two-channel system, UHJ. ("U" for Universal, "H" for Matrix H (hearing), a BBC system whose patents are now part of the Ambisonic pool, and "J" for 45J, the original two-channel Ambisonic system. It is sometimes called BHJ to acknowledge the Cooper-Shiga BMX system, which is also part of this pool.) This system audibly outperforms *any* four-channel discrete system.



Reprinted with the permission of IMF Electronics.

FIGURE 9: Overall UHJ decoder circuitry. In a VDP decoder, the $-jW$ *signal would be dynamically varied by logic circuitry to control the direction of lowest phasiness.*

Two Is Greater Than Four

I have already asked you to swallow quite a bit by saying that a three-channel system is superior to any discrete technique. A careful examination of the psychoacoustics involved and a brief listen will confirm this. But to say that a mere two channels can provide exceptional performance really strains credulity. After all, simple two-channel matrices were awful, having to await such developments as the QS Vario-Matrix and the Tate Directional Enhancement System for SQ. Certainly, UHJ must need some kind of logic enhancement? No way! It beats SQ and QS (not to mention discrete) at their own game with simple, linear circuitry (*Fig. 9*).

Before I explain how this is possible, I ought to describe what you can expect to hear from UHJ playback. It is so good that many listeners will not be able to tell it from the full three-channel, B-format system. The sound field is equally coherent. You have no sense of four distinct sound sources, but instead perceive a continuous field. Moving close to one speaker does not degrade the image as badly as it does in quad or regular stereo. In addition, it is possible to turn sideways and hear the front sounds to one side, the rear to the other. As you sidle up to the front, the front sounds predominate, and vice-versa. The listening window is wider and deeper than that for regular stereo, even extending outside the speaker boundaries.

These effects are not possible with discrete quadraphonics, let alone matrix systems. A minute or so of listening will sell even the most jaded audiophile on two-channel Ambisonic sound. Why, then, outside of up/down effects, do we even consider three-channel Ambisonic systems? The reason has to do with the inevitable trade-offs involved in reducing the signal from three to two channels.

The W, X and Y signals are mixed to produce a compatible stereo signal. (The stereo is actually slightly better than average, since sounds at the far left or right will appear outside the speakers.) UHJ is a "kernel" system, so the encoding for any position is exactly specified. After dematrixing (to produce approximations to the original W, X and Y signals), the signal must again go through the speaker-feed matrix. The amazing thing is that it is possible to arrange this processing so that the *exact*

Makita localization is retained for *all* directions. This is why UHJ has such remarkably good performance.

The trade-off is a significant increase in phasiness. Sounds take on a slightly bloated quality and come forward from the speakers. A subjective increase in coloration, caused by the quadrature phasiness components' altering of the frequency response, is also noticeable. Fortunately, there is a solution for this.

The ear and brain are much more sensitive to phasiness to the front than to the rear. You can adjust the decoding so that there is virtually no phasiness in the front, at the expense of high phasiness to the rear. For a forward-facing listener, however, the effect is one of a large improvement at the front, with little degradation to the rear. (Of course, if you turn around, you can hear the phasiness.) It is possible to dynamically alter the phasiness reduction so that the loudest sound is the least phasey. A special decoder using this Variable Directional Preference (VDP) technique might be available soon, but such a system enhances full surround recordings the most. VDP produces little improvement with ambience-only recordings, and none is really needed.

Optimized Two-Channel Encoding

You can see the advantages of UHJ encoding by manipulating the encode/decode equations, plugging the results into the speaker-feed equations, and relating all this to the psychoacoustic localization formulas mentioned earlier. But there are other ways of explaining what happens in a two-channel surround system. The Scheiber Sphere is one option.

In any two-channel surround-sound system, we are interested in three parameters—the intended direction of the sound source, the relative amplitude of the signals in the two channels and the relative phase of the channels. (The last two parameters, of course, represent the way we encode the first.) The usual way of representing the relationships of these parameters graphically is by plotting the encoding locus on the surface of a sphere. In the US, it is

The Psychoacoustics of Directional Hearing

The following approach considers only sounds originating in the horizontal plane, although you may extend the methods to periphonic (with-height) systems. Consider the X and Y axes as pointing forward and to the left, respectively. Assume that N loudspeakers are situated on a circle in the azimuthal directions phi-sub-i (i = 1,2,...N) measured counterclockwise from the X axis (due front). For simplicity, assume that all sources lie at a long distance from the center so that sounds arrive as plane waves. If a given mono sound is fed to all speakers, with the complex gain phi-sub-i to the i'th speaker, the following parameters influence the localization of that sound.

1. *Makita's localization* describes the localization azimuth for listeners facing the apparent sound source at low frequencies (less than 1kHz). Calculate x and y as follows:

$$x = \mathrm{Re}\left[\frac{(\Sigma P_i)(\cos\phi_i)}{\Sigma P_i}\right]$$

$$y = \mathrm{Re}\left[\frac{(\Sigma P_i)(\sin\phi_i)}{\Sigma P_i}\right]$$

where Re means "real part of," and the sums are over i = 1 to N.

Makita's localization azimuth (θ_v) is given by the following equation:

$$x = (r_v)(\cos\theta_v)$$
$$y = (r_v)(\sin\theta_v)$$

where the velocity vector magnitude (r_v) is greater than zero.

2. The *velocity vector magnitude* (r_v) equals

$$\sqrt{x^2 + y^2}$$

where x and y are as defined above. It describes the stability of sound localization with head rotation at low frequencies. For "natural" sounds, r_v equals 1. If r_v is much greater than 1 for a reproduced sound, you can hear an out-of-phase effect. If r_v is close to zero, you hear an "in-the-head" or "close-to-the-head" effect, along with excessive image movement when you move your head. (This effect is under complete control in Ambisonics, allowing the producer to move the sound toward or away from the listener.)

3. *Phasiness* (q) most often affects forward-facing listeners. It is an unpleasant "pressure-on-the-ears" sensation and might include image blurring. Its value is given by the following formula:

$$q = \mathrm{Im}\left[\frac{(\Sigma P_i)(\sin\phi_i)}{\Sigma P_i}\right]$$

where for real u, v, Im(u + iv) means v. Ideally, for natural sounds, q should equal zero, but |q| less than 0.21 is relatively innocuous, |q| less than 0.5 is generally tolerable, and |q| greater than 1 is unacceptable.

4. The *energy vector azimuth* describes the image azimuth at high frequencies (500Hz to 5kHz) and also appears at low frequencies for slightly off-center listeners. Calculate it as follows:

$$x_E = \frac{(\Sigma|P_i|^2)(\cos\phi_i)}{\Sigma|P_i|^2}$$

$$y_E = \frac{(\Sigma|P_i|^2)(\sin\phi_i)}{\Sigma|P_i|^2}$$

where

$$x_E = (r_E)(\cos\theta_E)$$
$$y_E = (r_E)(\sin\theta_E)$$

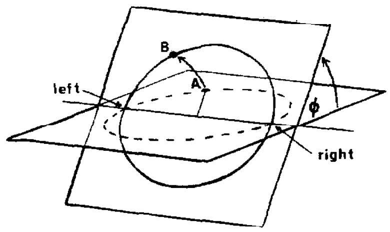
and r_E is the energy vector magnitude. This describes the apparent localization, especially when the listener faces the apparent direction.

5. The *energy vector magnitude* (r_E) equals

$$\sqrt{x_E^2 + y_E^2}$$

where x and y equal the values defined above. This quantity describes the stability of the sound image with head movement, especially at frequencies between about 500 and 5,000Hz. Ideally for "natural" sounds, r_E equals 1, and for reproduced sounds r_E can never exceed 1. In practice, an r_E of 0.7 is excellent, 0.5 is quite acceptable, and 0.35 is tolerable for sounds to the rear of the listener.

When the encode/decode equations for B-format Ambisonics are used to derive speaker-feed signals and these signals are plugged into the directional equations given above, all the equations are solved *exactly* for all directions. □



Reprinted, with permission, from the JAES *(Volume 23, Number 2), March 1975, p. 117.*

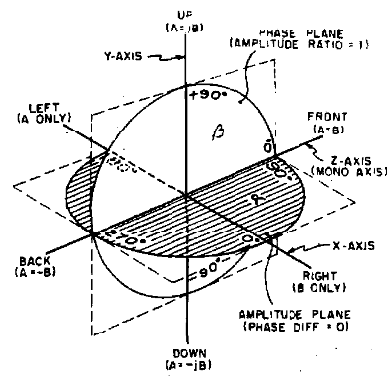
FIGURE 10a: Representation of a stereo position on a sphere. Point A in the horizontal plane represents a sound recorded with a gain of α on L and β on R. Point B represents a sound recorded with α (phase shifted ϕ) on L and with β on R.

called the Scheiber Sphere, after the first person to describe the concept. The British, in dogged refusal to acknowledge Peter Scheiber's contributions to surround sound, insist on calling it the Energy Sphere or a Stokes-Poincaire plot.

The plotting goes like this. The azimuth (that is, the angular position in the horizontal plane) equals [2 × arctan(L/R)], where L and R are the amplitude (including signal polarity) of the left and right transmission channels. Zero degrees is considered to be due right on the sphere. A signal on only the right channel will appear due right. Signals of equal strength on both channels will point forward (90 degrees), while signals of equal strength but opposite phase will appear at −90 degrees. A left-only signal will appear due left at 180 degrees. If this is not clear, work through the trig on your calculator.

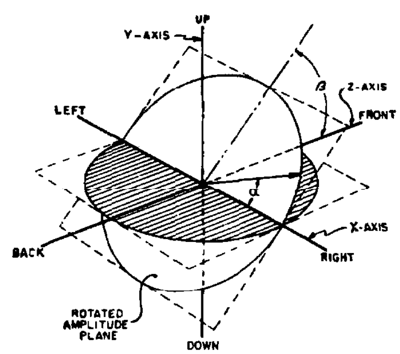
Phase differences between the channels are shown by angles of elevation and depression. L is considered to "lead" R, so if L's phase is ahead of R's, the point is plotted above the "equator," and vice-versa. Angles vary from 0 degrees (at the equator) to ± 90 degrees at the poles. See *Figs. 10–12.*

As I mentioned before, do not confuse polarity inversion with phase shift. If one of the encoded channels is of opposite polarity with respect to the other, then its position along the locus will be *on* the equator, not above it, since the calculation of azimuth includes signal polarity. [For



Reprinted, with permission, from the JAES *(Volume 19, Number 9), November 1971, p. 113.*

FIGURE 10b: Spherical α,β (amplitude ratio, phase difference) coordinates showing amplitude and phase planes. The positive Z axis is designated the mono axis because the level with which any matrixed channel of information is reproduced in monaural playback depends on the proximity of its α,β coordinates to those of the mono axis (90.0).

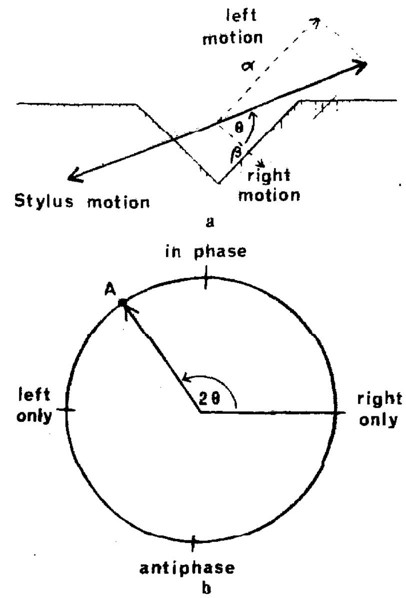


Reprinted, with permission, from the JAES *(Volume 19, Number 9), November 1971, p. 113.*

FIGURE 10c: The plane of the amplitude ratio angle, rotated by the plane difference parameter β.

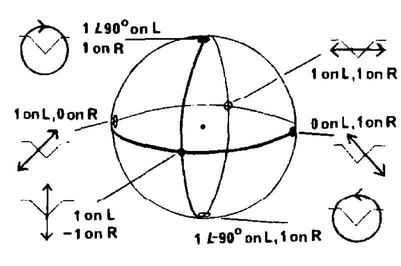
the arctan of an angle to fall between 90 and 270 degrees, either the sine or cosine of the angle (but not both) must be negative.]

Most people immediately grasp these concepts, then get bogged down in a very common and understandable mistake. They think that the azimuth of the locus necessarily corresponds to the intended direction of the sound source. It *may,* but it does not *have* to. (As we will see later, RM shows exact correspon-



Reprinted, with permission, from the JAES *(Volume 23, Number 2), March 1975, p. 117.*

FIGURE 11: Point A at an angle 2θ from the right point of b represents a direction of stylus motion at an angle θ from the right wall's direction of motion as in a.

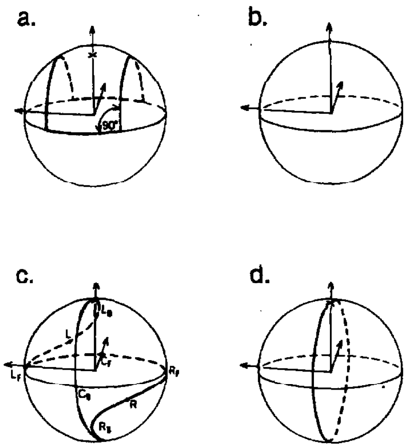


Reprinted, with permission, from the JAES *(Volume 23, Number 2), March 1975, p. 117.*

FIGURE 12: Stereo positions and stylus motions corresponding to various points on the energy sphere.

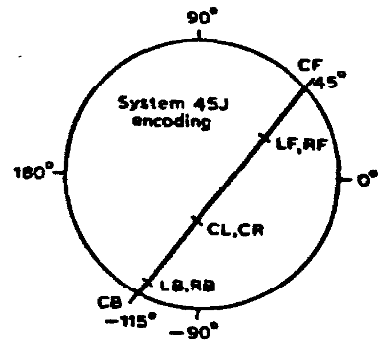
dence, whereas UHJ does not.) To avoid confusion, the intended direction should be marked on the locus. A simple example will illustrate this point.

Forget about surround sound for a moment, and suppose that we have plotted regular stereo on the sphere. A source that is on only the left channel will be plotted due left on the sphere (180 degrees). Our intuition suggests that this represents due left (or center-left). But in the listening



Figures a, c and d reprinted, with permission, from Nature, *December 13, 1974. Copyright © 1974 MacMillan Journals Limited.*

FIGURE 13a: Scheiber Sphere plots of some representative horizontal pan-loci: a is the Sansui QS, b is the RM, c is the CBS SQ, and d is a great-circle locus consistent with Ambisonic encoding. In sphere c, note the cusps and left-right asymmetry due to the choice of front-sector mapping and the limitations of encoding from four pair-wise blended channels. CF is the center front, CB is the center back, L is full left, R is full right, and LF is left front.



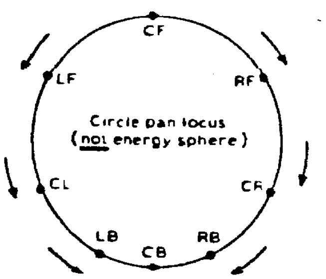
Reprinted, with permission, from Wireless World, *April 1977, p. 38.*

FIGURE 13b: Scheiber Sphere plot of two-channel UHJ encoding.

room, this signal will appear at the left speaker, which is usually about 30 degrees left of center-front, *not* due left.

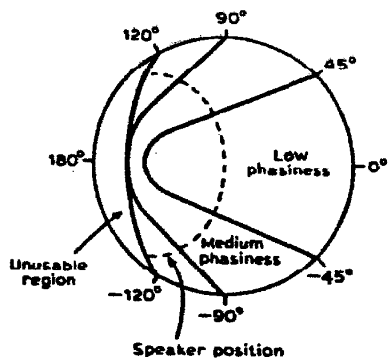
Let's examine the Scheiber Sphere plots for RM, QS, SQ and UHJ *(Figs. 13a* and *13b).* One thing not immediately obvious is that RM and UHJ are kernel systems. Their system-encoding specifications describe *exactly* how each direction is to be encoded. On the other hand, QS and SQ are matrices, which are specified for only the four corner positions. The intermediate positions shown have been computed by assuming PWM between adjacent channels. When Sansui says that QS is "derived from" RM, it is true only in that the four corner positions have the same encoding in each system. QS assumes, like all other matrix/discrete systems, that signals are to be pan-potted between the channels.

It should be obvious that SQ and QS have serious problems. The worst of these is that the encoding loci are not smooth. As you can see, both have distinct cusps, with SQ the worse of the two. This means that points that are relatively far apart in "real" space will be relatively close together in "encoding" space. That's a no-no! Both SQ and QS depend on sophisticated "logic" decoders for acceptable performance. These en-



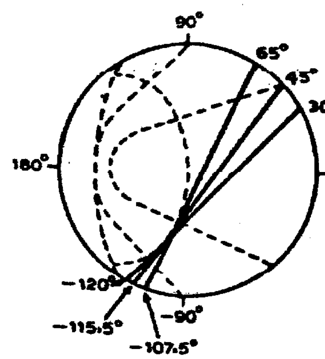
Reprinted, with permission, from Wireless World, *April 1977, p. 38.*

FIGURE 14: This circle shows the optimized nonsymmetric distribution of different encoded directions within the circle "pan-locus." CF is center front, RB is right back, CL is center left, and so on. This is not a picture of the energy sphere.



Reprinted, with permission, from Wireless World, *April 1977, p. 38.*

FIGURE 15a: The quality of mono and stereo reproduction shown on the energy sphere as viewed from the right side. The speaker position curve indicates that the signal appears to come from only one speaker in stereo.



Reprinted, with permission, from Wireless World, *April 1977, p. 38.*

FIGURE 15b: Three possible choices of the two-channel encoding system having optimized mono and stereo reproduction.

coding cusps make it more difficult for the logic circuitry to determine the intended directionality. When UHJ decoders use logic circuitry to reduce phasiness, this kind of problem cannot occur because UHJ has a smooth locus.

Another problem with SQ is its lack of symmetry. Some abstruse mathematics in Peter Fellgett's article "Surround Sound '76" (*Hi-Fi Sound Annual 1976*) shows that a nonsymmetrical matrix cannot be decoded (that is, manipulated to produce speaker-feed signals) in such a way that all the psychoacoustic criteria governing low-frequency directionality are satisfied simultaneously. Before the QS supporters start snickering, it turns out that the side signals in QS cannot be decoded correctly either. Although both systems are bad, SQ is worse than QS.

Both RM and UHJ allow correct decoding for all signals. Why doesn't UHJ use the same encoding as RM? Why does the locus have that odd tilt to it? And why are the sources spread out all over the place? Let's look at the last question first.

The ear and brain do a much better job of localizing sounds to the front than to the rear. Logic suggests we arrange the encoding so that frontal sounds take up as much of the encoding space as possible. This ensures that any transmission or decoding errors will have minimal effect on the perceived positions of frontal sounds. Of course, the rear sounds then get the worst of it, but it does not matter because the brain does not localize them as well anyway. This is the philosophy of UHJ, where the front half-circle of "real" space gets 240 degrees of the encoding locus. RM is fully symmetrical, with front and back taking up equal encoding space (*Fig. 14*).

Now for the tilt. It was introduced to improve mono and stereo compatibility and to reduce the phasiness of rear sources as heard by stereo listeners. Let's examine compatibility first (*Figs. 15a* and *15b*).

Compatibility problems are not new: they arrived with stereo. The most common is the build-up of the center signal when a stereo recording is played in mono. The channels add, giving a 6dB boost to a central performer. Sounds not quite center are increased proportionately less, while sources to the far left or right get no increase at all. The soloist becomes unduly prominent, a problem with all types of music. When matrix quad came in, a new incompatibility appeared. Sounds at or near center-back were recorded anti-phase (QS, RM, SQ), which caused them to cancel or be severely attenuated in mono. (Stereo listeners also perceive phasiness and lack of localization.)

Therefore, the word went out: thou shalt not encode center-rear signals in matrix quad. But why should the producer be prohibited from placing the sounds to the rear so that *mono* listeners will not miss anything, when the *raison d'etre* of surround sound is the ability to place sounds *anywhere* around the listener?

One goal of Ambisonics was to allow *any* type of recording to be produced, without favoring one technique over another. This means that a record producer should be free to place the performers *anywhere* around the listener, without losing anything in mono or producing irritating aural quirks in stereo. That's what compatibility is all about.

Ambisonics solves the compatibility problem (as well as reducing phasiness for rear signals) by the simple expedient of tilting the encoding locus. To see how this works, look at *Fig. 15a.* It shows the Scheiber Sphere from the right side. The encoding space has been marked off to show the degree of phasiness perceived by stereo listeners. As sources move closer to center-rear, the encoded signals are less and less in phase with each other, especially for encoding loci that are strictly on the equator, as in RM. Sounds encoded close to center-rear but not dead-on approach full inversion less closely and show correspondingly less phasiness. *Figure 15a* reflects phasiness caused by polarity differences, as well as interchannel phase shift. Notice how, as the angle of elevation (i.e., interchannel phase shift) increases, the area of unacceptable phasiness moves closer to the front of the sphere.

Lowering and tilting the locus gives us what we need. (See *Fig. 15b* for the variety of possible center-front phase shifts.) All rear sources are now moved out of the positions of unacceptable phasiness. This comes at the expense of *increased* phasiness for most frontal signals, but, paradoxically, this is desirable! For one thing, the extra phase shift causes sounds encoded at the sides to move out beyond the loudspeakers.

We also get the desired enhancement of compatibility. Center-front sounds are now 45 degrees out of phase, limiting their mono build-up to only 3dB. (The maximum amount of stereo phase shift tolerable to critical listeners is 45 degrees.) With the shift reduced to 115 degrees or less, center-rear signals are no longer anti-phase, giving a maximum mono attenuation of less than 2dB.

Going Around in Circles

You will notice that the encoding locus for UHJ is a circle. Is this optimum? Yes. We have seen that a smooth locus, free of kinks and cusps, is desirable, and a circle meets this requirement perfectly. Furthermore, the circle should be a "great" circle—that is, its center should pass through the center of a sphere. A great circle has the largest possible circumference of any circle that can be drawn on a sphere, so we are assured of using the maximum amount of encoding space. The UHJ locus sits a bit below center, so it is not *quite* a great circle. This allows the encoding locus to touch the locus of apparent single-speaker sources, which is required to ensure a good spread of sound between the speakers.

Why is a great circle optimum? Why not use an elongated, twining locus, like the stitching on a baseball? Wouldn't that use much more of the available encoding space? Yes, it would, but it would be worse in other respects. Notice that walking around the sound field is the same thing as traversing the encoding locus. If you follow this on your baseball, you will see that encoding points that are well separated along the length of the locus are nonetheless relatively close together on the surface of the sphere (i.e., close in encoding space). A "serpentine" locus would, therefore, defeat the reason for enlarging the encoding space.

One other point is worth addressing. Some of you might be wondering whether Ambisonics is "holographic." (Its supporters sometimes describe it as "quasi-holographic" because its performance is so remarkable.) An optical holograph presents to the eye essentially the same waveform it would receive if it were actually looking at the object. For a comparable sonic result, it would be necessary to duplicate the sound field over a very large listening area. Currently, this is impossible, so Ambisonics takes a different tack.

As we have seen, the most important parts of the sound field for determining directionality are the zero and first-order components. Ambisonics extracts these and presents them to the listener in a psychoacoustically optimum way. This produces accurate and stable imaging over a wide listening area. It is holographic only in the sense that it properly reproduces those parts of the sound field that are required for good localization. No other so-called system does this.

Next time, Mr. Sommerwerck will give construction and recording details for the Ambisonic system.