Direction and Space, the Final Frontiers

How Many Channels Do We Need To Be Able To Believe That We Are "There"?

by

Floyd E. Toole, Ph.D., Vice President Acoustical Engineering
Harman International Industries, Inc.

The capture, storage and reproduction of musical sounds are the missions of the professional and consumer audio industries as we know them. Over the decades, we have struggled, artistically and technically, to improve the accuracy with which the sounds are recorded and reproduced. Accuracy of timbre ensures that the voices and instruments sound like the real thing. However, there is much more. Musical sounds do not exist in isolation, they exist in acoustical contexts, either real or synthesized: concert halls, jazz clubs, and so on. Ideally, listeners should be able to close their eyes and believe that they are "there", with the performers. We are much closer to our goal than ever before.

In the beginning, there was mono. Everything we heard was stored in and reproduced from a single channel. In those early days, listeners enthused, and critics applauded the efforts of Edison, Berliner, and others, as being the closest possible to reality. They were wrong, but clearly a revolution in home entertainment had taken place.

With two-channel stereo came dramatic improvements in the impressions of direction and space. Once we got past the exaggerated "ping-pong", "hole-in-the-middle" problems of many early recordings, listeners enthused, and critics applauded the efforts of many artists and recording engineers as being the closest possible to reality. They were wrong again, but clearly another revolution in home entertainment had taken place.

Now we demand still more . . . more realism, more dramatic effects, and more listeners to share the auditory experiences. Multichannel audio is now a reality. Is this the solution that we have been searching for?

In this two-part presentation we examine how well we are doing in meeting our objectives. We take a look at some of the history, review the present, and speculate about the future.

PART ONE: THE MULTICHANNEL OPTIONS

A BRIEF HISTORY

Monophonic[1] reproduction conveys most of the musically important dimensions: melody, timbre, tempo and reverberation, but no sense of spatial envelopment - of being there. In 1930's the essential principles by which the missing directional and spatial elements could be communicated were understood, but there were technical and cost limitations to what was practical. It is humbling to read the wisdom embodied in the Blumlein-EMI patent [1] applied for in 1931 describing two-channel stereo techniques that would wait 25 years before being exposed to the public. Then there are the insights of the Bell Telephone Laboratories [2] who, when considering the reproduction of auditory perspective, concluded in 1934 that there were two alternative reproduction methods that would work: binaural and multichannel.

By binaural, they meant the technique of capturing a multi-dimensional sound field using microphones at the ear locations in an artificial head (thereby encoding all of the directional cues in the left and right-ear signals) and reproducing those signals through headphones. The listener's ears would then "hear" what the dummy head "heard" and, in theory, perfect auditory perspective would be communicated. Binaural reproduction is the only true justification for the "we have two ears, therefore we need two channels" argument. Two channel stereo, as we have known it, is the simplest form of multichannel reproduction – it is not binaural.

Multichannel reproduction is more obvious, since each channel and its associated loudspeaker creates an independently localizable sound source, and interactions between them create even more. Inevitably, the question arose, "how many channels are necessary?" Bell Labs scientists concluded that a great many channels would be necessary to capture and reproduce the directional and spatial complexities of a musical soundstage – not even attempting to recreate a surrounding sense of envelopment. Being practical, they investigated the possibilities of simplification, and concluded that, while two channels could yield acceptable results for a solitary listener, three channels (left, center and right) would be a desirable minimum to establish the illusion of a stable front soundstage for a group of listeners.

By 1953 ideas were more developed and, in a paper entitled "Basic Principles of Stereophonic Sound" [3], William Snow describes a stereophonic system as one having two *or more* channels and loudspeakers. He says: "The number of channels will depend upon the size of the stage and listening rooms, and the precision in localization required." He goes on " . . . for a use such as rendition of music in the home, where economy is required and accurate placement of sources is not of great importance if the feeling of separation of sources is preserved, two-channel reproduction is of real importance."

---

[1] Monophonic describes reproduction through a single channel. Monaural means, literally, one ear. We listen binaurally (through two ears) no matter how many channels are used.

So, two-channels were understood to be a compromise, “good enough for the home” or words to that effect, and what did we end up with? Two channels! The choice had nothing to do with scientific ideals, but with technical reality that, at the time stereo was commercialized, nobody knew how to store more than two channels in the groove of an LP disc.

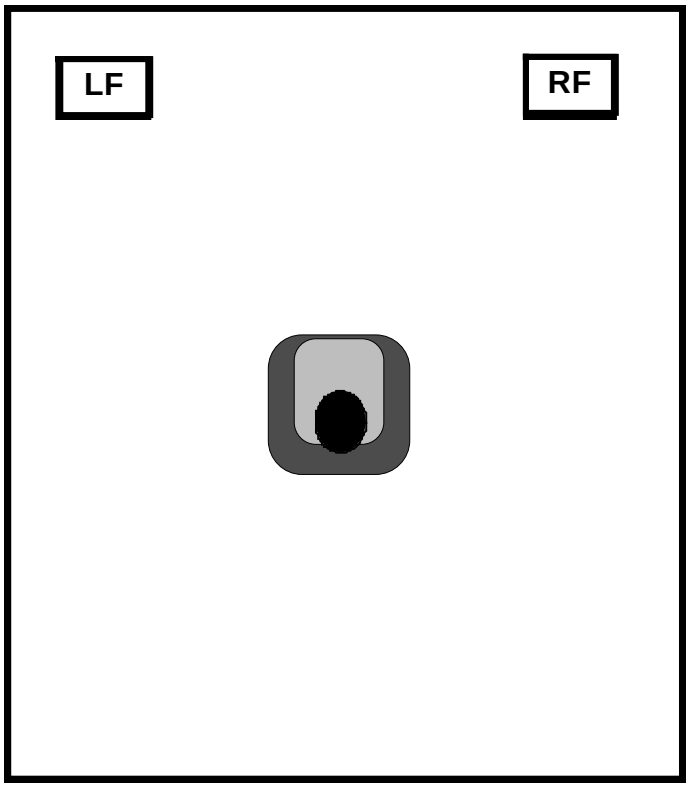
However, around that same time, the film industry managed to succeed where the audio industry failed, and several major films were released with multichannel surround sound to accompany their panoramic images. These were discrete channels recorded on magnetic stripes added to the film.

Although they were very successful from the artistic point of view, the technology suffered because of the high costs of production and duplication. Films reverted to monophonic optical sound tracks, at least until the development of the "dual bilateral light valve". This allowed each side of the optical sound track to be independently modulated, and two channels were possible. As we will see, it didn’t stay that way for long and, ironically, it has been the film industry, not the audio industry, or audiophiles, that has driven the introduction of multichannel sound reproduction in homes.

MULTICHANNEL SOUND – A HUMBLE BEGINNING

The arrival of two-channel stereo in the ‘50’s was a revolution, even though recording techniques being used at the time frequently resulted in hole-in-the-middle sound stages, and exaggerated left-right effects. Stereo is not blessed with an underlying encode/decode system or philosophy - it is merely a two-channel delivery mechanism. Over the years the struggle to capture, store and reproduce realistic senses of direction and space from two channels and loudspeakers has been a mighty one. There has been no single perfectly satisfactory solution, even after all these years. Professional audio engineers have experimented with many variations of microphone types and techniques, trying to capture the directional and spatial essence of musical events. At the playback end, multitudes of signal processors, loudspeaker designs and “tweaks” have come and gone over the years, all attempting to extract a more gratifying sense of space and envelopment. Stereo truly is a “spatially-deprived medium”.

What can one say about a system that accommodates loudspeakers having directional characteristics ranging from omni-directional, through bi-directional in-phase (so-called “bipole”), bi-directional out-of-phase (dipole), predominantly backward firing, and predominantly forward firing? The nature of the direct and reflected sounds arriving at the listeners’ ears from these different designs runs the entire gamut of possibilities. This is not a system at all, it is merely a foundation for individual experimentation. The history of two-channel stereo is littered with examples of efforts to generate a more engaging sense of envelopment and depth, some acoustical, some electronic, and some that appear to operate simply on faith. Remember the Hafler system [4]sold by Dynaco? How about Carver’s Sonic Hologram?[5] Nowadays we have SRS, Spatializer and hosts of “DSP’s” adding dimensional embellishments. We can only conclude that, as a multichannel system, two channels are simply not enough.



Added to these fundamental problems is the inconvenience of the “stereo seat”. Because of the stereo seat, two-channel stereo is an antisocial system - only one listener can hear it the way it was created. If one leans a little to the left or right, the featured artist flops into the left or right loudspeaker and the soundstage distorts. When we sit up straight, the featured artist floats as a phantom image between the loudspeakers. This puts the sound image where it belongs in space, but then there is another problem, the sound quality is altered because of the acoustical crosstalk.

In stereo systems, the sound from each loudspeaker travels to both ears. When identical sounds radiate from both channels, as happens for a center image, there is a comb-filter effect at each ear when the direct sound from the nearer loudspeaker acoustically interferes with the slightly delayed sound from the opposite loudspeaker. The dominant effect is a distortion of the amplitude and phase response of the sound of the center image. Ironically, no matter how perfect a loudspeaker may be in creating a flat frequency response and a linear phase, those features will not be appreciated in the sound of the center image because of a limitation of two-channel stereo itself. You don’t believe me? Play some monophonic pink noise and move in and out of the stereo sweet spot. As you move from the left or right towards center you will experience phasiness, and as you approach the precise center location, the sound will get noticeably duller as the destructive interference dip around 3-4 kHz develops. Fortunately, room reflections help to minimize the annoyance of the effect in most home installations.

In fairness it must be said that, after over forty years of experimentation, the best two-channel stereo recordings reproduced over the right set of loudspeakers in the right room can be very satisfying indeed. Sadly, only a fraction of our listening experiences fall into that category. Clearly stereo does not get us to our long-term objective.

MULTICHANNEL SOUND – TWO PLUS TWO EQUALS QUAD

In the ‘70’s we broke the two-channel doldrums with a misadventure into four-channel, called quadraphonics. The intentions were laudable - to deliver an enriched sense of direction and space. The key to achieving this was in the ability to store four channels of information in the existing two channels, and then to recover them.

There were two categories of systems in use at the time, matrixed and discrete. The matrixed systems crammed four signals into the bandwidth normally used for two channels. In doing this, something has to be compromised and, as a result, all of the channels did not have equal channel separation. In other words, information that was supposed to be only in one channel, would appear in smaller quantities in some or all of the other channels. The result of this "crosstalk" is confusion about where the sound is coming from. I well remember feeling as though my head was inside a cello in one of the quadraphonic recordings in my collection.

Various forms of signal-adaptive "steering" were devised to assist the directional illusions during the playback process. The “alphabet soup” is memorable: SQ from CBS, QS from Sansui, E-V from ElectroVoice and others. Peter Scheiber, a musician with a technological bent, figures prominently as a pioneer in the matrix game, with his patented encoder and decoder ideas being incorporated into many of the designs. The best of these systems were remarkably good in creating the illusion of completely separate, or discrete, channels when an image was panned around the room. However, this clear separation breaks down when there is a demand for several simultaneously-occurring discrete images.

Ultimately, there is nothing quite like four entirely discrete channels. However, achieving this on the vinyl LP's required that the recorded bandwidth be extended to about 50 kHz - quite a challenge. Nevertheless, it was accomplished, as CD-4 from JVC, and although this quadraphonic format did not survive, the technology necessary to achieve the expanded bandwidth did have a lasting benefit on the quality of conventional two-channel LP's. Half-speed cutting processes, better pressings and playback cartridges with wider bandwidth and reduced tracing and tracking distortions were to live on.

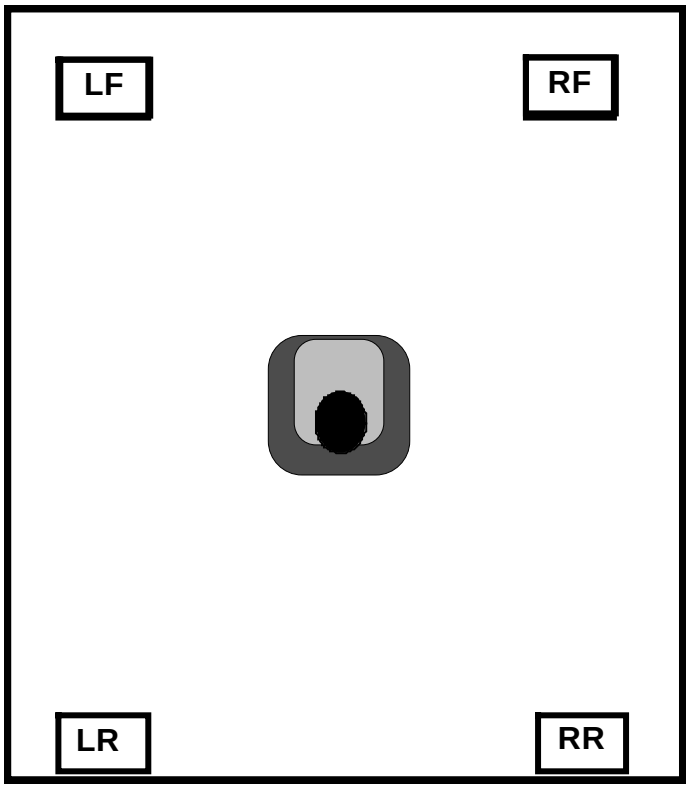
Discrete multichannel tape recordings were available, but open reel tape was a nuisance to say the least, and high-quality packaged tape formats (e.g. cassettes) were not yet ready for true high fidelity multichannel sound

Years passed, with the industry unable to agree on a single standard. Eventually, the whole thing dissolved into competitive squabbles. The industry lost a lot of money and credibility, and customers were justifiably disconcerted.

Looking back on this unfortunate episode in the history of audio, one can see a another reason for failure - the system was not psychoacoustically well founded. Lacking an underlying encode/decode rationale, the problems of two-channel stereo were simply compounded. There were even naïve notions of “panning” images front to back using conventional amplitude-panning techniques. The quadraphonic square array of left and right, front and rear, was still an antisocial, system, with even stricter rules. The sweet spot now was constrained in the front-back as well as the left-right direction.

In addition, there was no center channel, a basic requirement if one wishes to eliminate the stereo seat. Placing the additional channels behind the listener is less than optimum for generating envelopment and a sense of spaciousness. Placement to the side is better. Sounds arriving from the rear are extremely rare in the standard repertoire of

music, but the need for a credible spatial impression is common. Ironically, a 1971 publication entitled “Subjective Assessment of Multichannel Reproduction” [6] showed that listeners preferred surround loudspeakers positioned to the sides, relative to ones placed behind, granting subjective rating scores that were two to four times higher. It seems as though nobody with any influence read it.

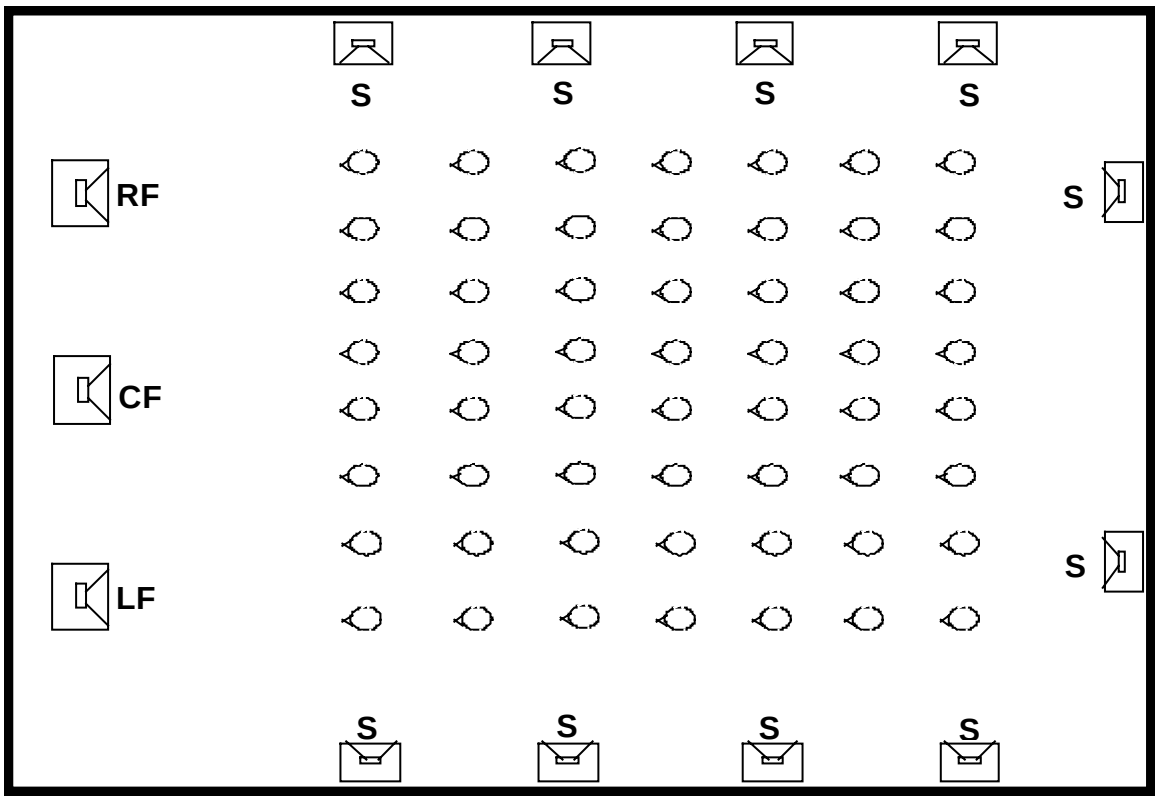


Fortunately, much of the innovation that went into quadraphonics was not wasted, it would live on in different forms.

MULTICHANNEL AUDIO - CINEMA MAKES IT WORK

Failure in one market, however, was not enough to kill good ideas, and here there were two: multiple channels and adaptive matrices. Dolby was well connected to the real multichannel pioneers, the movie makers, in the application of its noise reduction system to stereo optical sound tracks. Putting the pieces together, Dolby rearranged the channel configuration to one better suited to film use: left, center and right across the front, and a single surround channel which was used to drive several loudspeakers arranged beside and behind the audience. All of this was stored in two audio-bandwidth channels. With the appropriate adjustments to the encode matrix, and to the steering algorithm in the

active decoding matrix, they came up with the system that has become so familiar in quality films and cinemas: Dolby Surround or, as it is also known, Dolby Stereo.



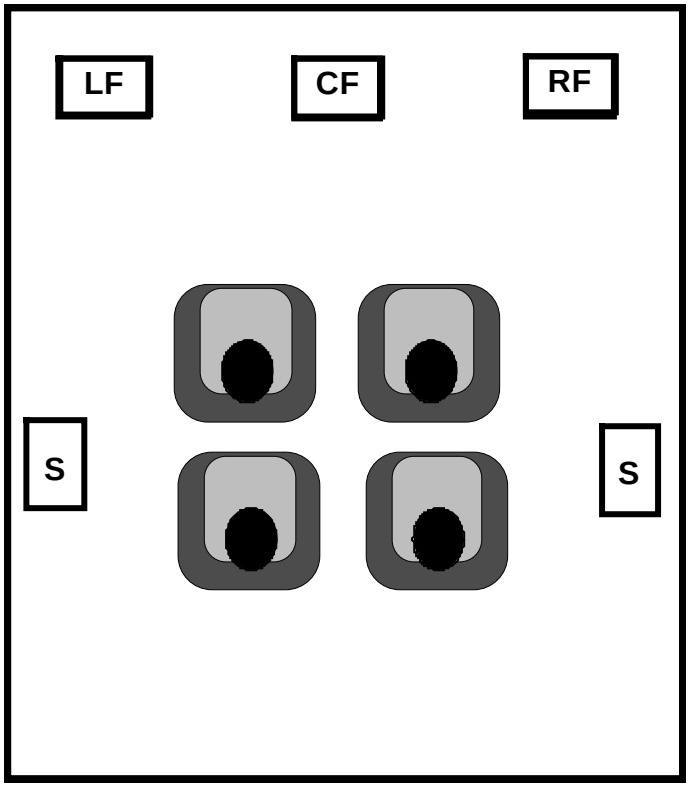
This system was subject to some basic rules that have set a standard for multichannel film sound: well placed dialogue in the center of the screen, music and sound effects across the front and in the surround channel. Reverberation and other ambience sounds are steered into the surround channel, as are sounds of aircraft, and the like. At times the audience can be enveloped in sound as in a football game, or it can be transported to a giant reverberant cave or gymnasium, or it can be inside the confines of a car engaged in a dramatic chase, or it can be treated to an intimately whispered conversation between lovers where the impression is that of being embarrassingly close. To achieve this dynamic range of spatial experiences requires a flexible multichannel system, controlled-directivity loudspeakers, as well as a degree of control over the acoustics of the playback environment. When it is done well, it is remarkably entertaining . . . and it is not antisocial! There are still better and worse seats, in the house, but there are several truly good seats. The basic format of a front soundstage with enveloping “ambiance” is also the basis for most of our real life musical experiences, so audiences were immediately comfortable.

It is important to note that the characteristics of the encoding matrix, the active decoding matrix, the spectral, directional and temporal properties of the loudspeakers and room (the cinema in this case) all are integral parts of the functioning of these systems. Fortunately, the film industry acknowledges the need for standardization and so, from the outset tried to ensure that sound dubbing stages, where film sound tracks are assembled, resemble cinemas, where audiences are to enjoy the results. Although the industry standards provided a basis, there were still inconsistencies. This left a need, and an

opportunity for Lucasfilm's THX to establish a program to certify the audio performance of cinemas, so that audiences would have an even greater assurance of quality.

MULTICHANNEL SOUND – MOVIE SOUND COMES HOME

With the popularity of watching movies at home, it wasn't long before Dolby Surround made its way there. Adapting it to the smaller environment required some changes, but nothing very radical. Reducing the number of surround loudspeakers to two, ensured greater consumer acceptance, and recommending the placement of these loudspeakers to the sides of the listeners ensured that they would be most effective in creating the required illusions of space and envelopment. Delaying the sounds to the surround speakers used the precedence effect to ensure that, even in a small room, the ambiguously localized surround sounds would be perceptually separated from those in the front channels.



At the outset, a simple fixed-matrix version was available in entry-level consumer systems. The fixed-matrix systems exhibited so much crosstalk among the channels that listeners were surrounded by sound most of the time, even when it was inappropriate.

As I recall, it was Fosgate and Shure HTS who brought the first active-matrix decoders to the home-theater market, albeit at premium prices. Low-cost silicon chips eventually brought active-matrix Dolby ProLogic decoding to the masses and home entertainment entered a new era. Admittedly it was audio for movies, but it was multichannel audio nevertheless, and many of us began to appreciate some of the

dimensions that were missing from our directionally- and spatially-deprived two-channel stereo lives.

Dolby Surround was designed for movie sound tracks reproduced in large cinemas, and that it does very well. However, once audiophiles get a taste of something attractive, they want more. In this case, the more they wanted was multichannel music.

Playing conventional stereo recordings through a Dolby ProLogic processor was a logical experiment, and most of us have done it. The results are spotty; some recordings work quite well, and others don't. A basic problem stems from the fact that material mixed for no center channel, when played through a conventional matrix decoder, yields center-channel signals that are perceived to be louder than they should be. The basis for this problem is in the translation from large cinemas to listening at shorter distances in smaller rooms. The high-frequency rolloff in the surround channel is also noticeable, and the active matrix steering is sometimes caught messing with the music. Recordings made specifically for Dolby Surround are better, but even they have failed to establish a significant following in the music recording industry. None of this is surprising, but all of it means that we have not yet arrived at a general-purpose multichannel solution.

THX EMBELLISHMENTS

In a natural succession to their THX program for certifying cinema sound systems, Lucasfilm established a licensing scheme for certain features intended to enhance, or in certain ways ensure, the performance of home theater systems based on Dolby ProLogic decoders. Home THX, as it was called, added features to a basic ProLogic processor and to the loudspeakers used in home theater systems, and set some minimum performance standards for the electronics and loudspeakers. At a time when the market was being inundated with "cheap and cheerful" add-on center/surround speakers and amplifiers, THX made a clear statement that that would not do; all channels had to meet the same standard. Tomlinson Holman deserves credit for assembling this amalgam of existing and novel features into what has become a benchmark for consumer home theater.

The THX embellishments of relevance to this discussion are:

1. High- and low-pass filters to approximate a proper crossover between the subwoofer and satellite loudspeakers. Elaborate systems did this anyway, but this brought an important feature to the mass market.
2. Electronic decorrelation between the left and right surround channels. Reducing the number of surround speakers to two, and putting them in a small room eliminated much of the acoustical decorrelation (randomization of the sounds arriving at the listeners' left and right ears) that multiple speakers in a large cinema accomplished automatically. Substituting electronic decorrelation is a good idea which was, to my knowledge, first introduced in the Shure HTS systems.
3. "Timbre matching" of the surround to the L,C.R (front) channels. In my view, this is a dubious feature. Sounds arriving from the sides, or even from random incidences, cannot and should not match the timbre of sounds arriving from the front. Timbre matching is not natural - the design of the external ears ensures that. However, it is a relatively minor matter in the larger scheme.
4. "Re-equalization" of the sound track to compensate for excessive treble that is usually built into film tracks to compensate for sound systems in large cinemas. A single

correction curve was chosen. This is a useful feature, but it needs to be an adjustable tone control, since sound tracks vary in treble balance – some can do with more treble cut and some seem to be fine with no correction.

In loudspeakers, THX required some control of the vertical dispersion from the L, C, R (front) units, and a bidirectional out-of-phase (an approximation to a dipole) configuration for the surrounds. Controlled vertical dispersion was intended to reduce the strength of the floor and ceiling reflected sounds from the front channels. Horizontal dispersion was not modified, which left the familiar wall reflections intact for those occasions when we want to play conventional stereo through the L and R loudspeakers.

The purpose of the bi-directional "dipole" surround loudspeakers was to increase the proportion of reflected sound that is generated by the two surround speakers, thereby compensating somewhat for the fact that there are only two of them.

Both of these are good ideas, but some real-world implementations have created a belief that, somehow, they were incompatible with the objectives of good music reproduction. Indeed, there have been some less than worthy examples of home theater loudspeakers, but one can easily say the same about conventional "music" loudspeakers. In principle there should be no reason to differentiate between them. Good design is good design.

THX certification is a licensing business, and certification of the components in an audio system is a manufacturer's option. It costs money, and it requires that certain features or functions be present, and it imposes some restrictions. From the consumer's perspective, it should be regarded as an assurance of a certain level of performance from a product. However, systems that are not THX approved, may also perform to the same or even higher standards of performance. You just have to take somebody else's word for it, or make up your own mind.

MULTICHANNEL SOUND - THE AMBISONICS ALTERNATIVE

There are two parts to the Ambisonics premise. The first is that, with the appropriate design of microphone, it would be possible to capture (record) the three-dimensional sound field existing at a point. The second part is that, with the appropriate electronic processing, it should be possible to reconstruct a facsimile of that sound field at a specified point within a square or circular arrangement of four or more loudspeakers. Therefore, this system distinguishes itself from all others so far, in that it is based on a specific encode/decode rationale. Several names are associated with the technology. The basic idea for this form of surround sound was patented first by Duane Cooper [7]. Patents were also granted to Peter Fellgett and Michael Gerson who were working simultaneously and independently in England. Peter Craven contributed to the microphone design and, with the support of the NRDC, the U.K. group commercialized the Ambisonics record/reproduction system.[8][9]

It is an enticing idea, and the spatial algebra tells us that it should work. And it does, up to a point. Ambisonics has enthusiastic supporters, but it remains a niche player in the surround-sound industry. Most people know little to nothing about it, although there are some encoded recordings [9], and the Soundfield microphone is used in some recordings. The scarcity of playback decoders is a clear problem. However, there are

other considerations that may be significant. Ambisonics is a system that requires special recordings and playback apparatus, that is incompatible with other multichannel systems (although it need not be), and that ends up entertaining a single listener. Mind you, that listener can be well entertained.

I have heard the system several times in different places (including a precise setup in an anechoic chamber), and I will admit that with large spacious classical works the system creates an attractively enveloping illusion for a listener with the discipline to find and stay in the small sweet spot. It tolerates a certain amount of moving around, but leaning too far forward results in a front bias, leaning too far backwards creates a rear bias, leaning too far left . . . well, you get the idea. Big, reverberant, recordings are more tolerant of listener movements of course. All of this should be no surprise in a system in which the mathematical solution applies only at a point in space, and then only if the setup is absolutely precise in its geometry, and the loudspeakers are closely matched in both amplitude and phase response. Room reflections absolutely corrupt the theory.

In order to reconstruct the directional sound intensity vectors at the center of the loudspeaker array, some amount of sound may be required to be delivered by many, or all, of the loudspeakers simultaneously – that is the way the system works. A practical problem then arises, because we listen through two ears, each at different points in space, and both attached to a significant acoustical obstacle, the head. If a head is inserted at the summation point, then it is not possible for sounds arriving from the right loudspeakers to reach the left ear without colossal head-shadowing effects, and vice versa. The system breaks down, and we hear something other than what was intended. What is heard can still be highly entertaining, and possibly greatly preferable to stereo, but it is not a "reconstruction" of the original acoustical event. For that, one would need to generate individual sound fields for each of the ears, in a way enhancing the crosstalk-cancelled binaural scheme described in Part Two.

There are numerous ways to encode and store the Ambisonics signals, and even more ways to process the signals into forms suitable for reproduction from different numbers of loudspeakers in different arrangements. All of these I have not heard. Ambisonics may yet play a significant role in our audio lives. Certainly having multiple digital discrete channels within which to store data can only be an advantage for it. As it has been promoted and demonstrated, however, there seems to be a lot of paraphernalia for just one listener.

MULTICHANNEL SOUND - MATRIX MANIPULATIONS, AND IMPROVING ON "NATURE"

Dolby Surround, and the corresponding playback algorithm Dolby ProLogic, have become industry standards for matrix encoding and decoding of multichannel material. Dolby Surround evolved in the days of optical sound tracks on films, with all of the noise and fidelity compromises that entailed. Consequently, features were included in the ProLogic decoders to deal with some specific idiosyncrasies such as "splashes" of sibilants and other high-frequency sounds leaking from the front channels into the surround channels, and causing unwanted distractions. A 7 kHz rolloff – a loss of high frequencies – in the surround channel is the most audible result of this feature.

None of this matters much in movie sound tracks, which were monitored through such a system while they were being made. However, consumers take a more liberal attitude, and they play stereo music through the decoder, only to find that the surrounds are a bit dull, and there may be a bit too much emphasis on the center channel – the stereo soundstage shrinks in width.

Recognizing an opportunity to improve on a good thing, inventors have had a field day manipulating the parameters of the matrixes, with time delays and with steering algorithms, all in attempts to finesse the multichannel decoders either to be more impressive when playing movies, or to be more compatible with stereo music, or both. The more adventurous ones have augmented the surround system with additional loudspeakers behind the listeners. Most of them allow for full bandwidth surround and, possibly, also rear channels.

The basic notion is to be able to make better use of a five-channel audio system, that may have been created for movie playback, for other sound sources. The emphasis on "home theater" applications of multichannel audio systems has turned off many consumers who might have an equal or even greater interest in the musical applications of the technology. Believe it or not, lots of stereo recordings can sound even better through a good two-channel-to-multichannel conversion algorithm. All of these systems provide consumers with new ways to get pleasure from their multichannel audio systems. Some purists frown on such meddling, especially with film soundtracks, but lots of people, including the author, find rewards in the artistry of some of these systems.

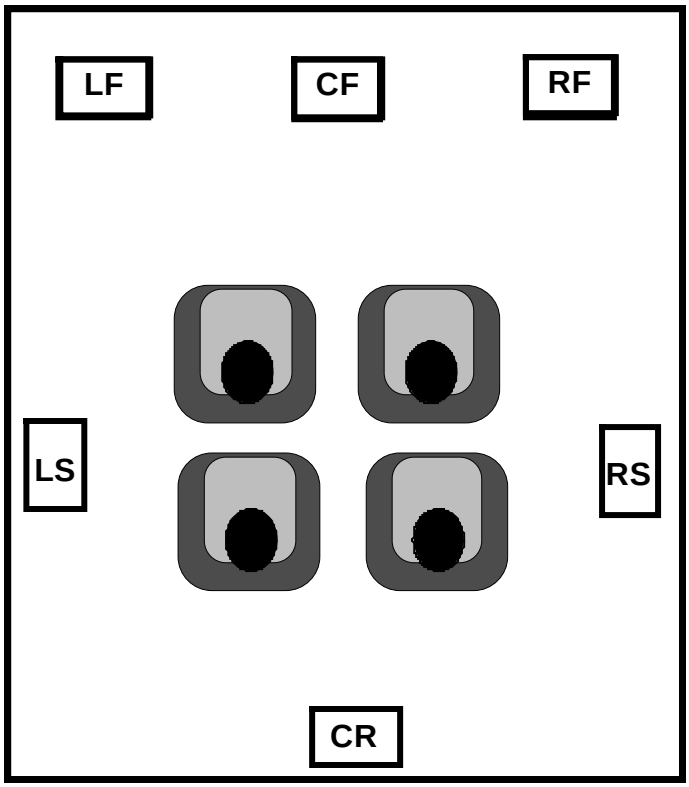
Just to be absolutely clear, when playing stereo music through these algorithms, we are hearing "ambience extraction", not reverberation synthesis. This is not the same as the often contrived and artificial sounding "hall" effects. All of the reverberation that is reproduced in the surround channels is reverberation that was in the recording. It is just redirected to the side and/or rear loudspeakers rather than being reproduced through the front channels. It sounds more natural. Occasionally, it can be a bit exaggerated, since the stereo recordings were not designed for this form of reproduction. In order to get a sense of spaciousness in stereo reproduction through two loudspeakers, more "ambiance" (decorrelated sound) has been recorded than would have been required if surround channels had been anticipated. The solution: use the remote control and turn the surrounds down a bit.

The 6-Axis algorithm

A veteran of the quadraphonic wars, Jim Fosgate found ways to decode Dolby Surround sound tracks in ways that many people found to be preferable to more conventional means. Part of the improvement had to do with the responsiveness of the steering logic, and part of it had to do with providing some amount of left-right distinction in the full-bandwidth surround channel. Since there is no such separation in the encoded program, the "art" has been to judge how much and when left and right front information should be directed to the surrounds, with what spectral modifications (if any), and with what delay.

Fosgate practiced his art well and, over the years, has generated several well received designs optimized for films and for different kinds of music, all in the analog domain. An interesting feature was the provision for separately driving the forward- and

rear- firing loudspeakers in the surround “dipoles” to allow for more directional and spatial enrichment. His designs can be found in products bearing his own name, as well as some older models from Harman/Kardon and JBL Synthesis, and current ones from Citation (citation-audio.com). Fosgate’s latest effort is called 6-Axis because, in addition to the basic five steered channels, there is an optional sixth, behind the listener to complete the surround effect.



The Logic 7 Algorithm

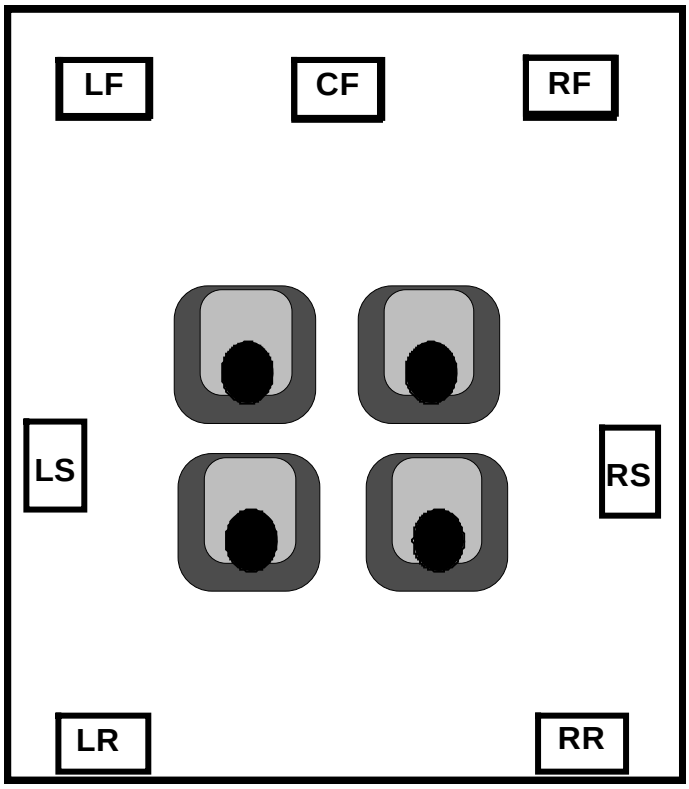
Working independently, and in the digital domain, Dr. David Griesinger, from Lexicon, has done similar things to move beyond the basic ProLogic process. Griesinger is probably best known in professional audio as the author of reverberation algorithms used in Lexicon products found in a high proportion of recording studios. Driven by an intense interest in the physics and psychoacoustics of concert hall acoustics, and a significant contributor to that area of science, it is no surprise that Griesinger’s efforts in surround-sound decoding and multichannel synthesis are based on his years of studying, synthesizing, and electronically enhancing the acoustics of concert halls. Accentuating the desirable aspects of complex multi-dimensional sound fields while avoiding undesirable artifacts is the essence of both endeavors.

The result is a suite of film and music playback algorithms embodied in Lexicon digital surround processors. Griesinger’s current product is called Logic 7, since it provides for two additional channels and loudspeakers behind the listener. Using a

sophisticated detection and steering process, these rear loudspeakers are supplied with strongly uncorrelated sounds, such as reverberation, applause and crowd sounds, or sounds that are strongly directed to move from front to surround, or vice versa. Thus the listeners (yes these are still very much social systems) are treated to a truly enveloping sense of ambiance, and to occasional sounds that sweep dramatically forwards or backwards, even with appropriate left or right biases. An important focus in the continuing development of Logic-7 is the quest for compatibility in multichannel reproduction of film sound tracks and music, as well as that between two-channel and multichannel reproduction of stereo music mixed for two channels.

Logic 7 decoders are in Lexicon's own surround processors (lexicon.com), JBL Synthesis processors, and in Harman Kardon receivers.

.



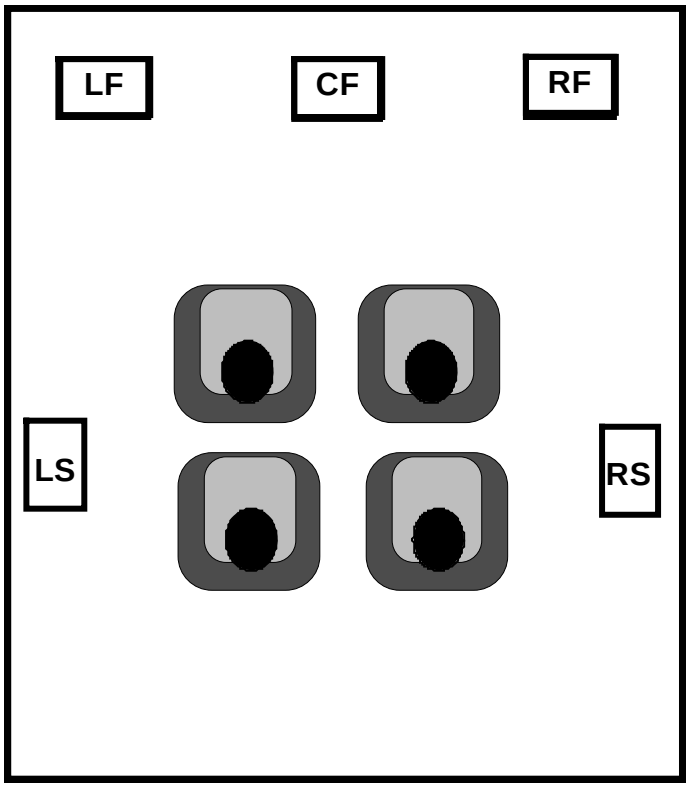
MULTICHANNEL AUDIO - DIGITAL AND DISCRETE

The few samples of discrete multichannel recordings from the quadraphonics era were sufficient to generate a lasting interest, if not an outright lust, to develop a viable format that did not suffer from leakage, or crosstalk, among the channels. Today we are experiencing several versions of that dream. Digital technology has given us an increasingly confusing collection of options, some originating in film sound, and others emerging from the music side of the business. There are the expected pro and con arguments about which ones sound better, but under the bluster and ballyhoo, all of the systems, so far, have sufficient sonic integrity that our entertainment is unlikely to be

compromised. Purist audiophiles have pushed for systems of such bandwidth and dynamic range that even the most fastidious super humans, dogs and aliens will be pleased. One of the wonders of digital encoding and decoding is that it is all possible for a price – bandwidth.

Currently established as a *de facto* standard for film sound, is Dolby Digital, also referred to as AC-3. This is a 5.1 channel system that was designed for movie sound tracks on film media. It is a "lossy" data reduction system, which uses clever psychoacoustics to simplify the data that are stored so that less digital space is required. The original data rate was determined by the limitation of how much digital data could be stored in the small spaces between the sprockets in a 35 mm film. However, versions exist with higher data-rates. A consumer version is used on laserdiscs, DVD's, satellite TV, and other media. Following the basic geometry of the existing multichannel system, Dolby Digital incorporates five channels, with completely separate left and right surround channels. All five channels are full bandwidth, discrete and equal-power, offering the multichannel producers enormous flexibility. A sixth channel is used for occasional very powerful low-frequency sound effects. Thus we end up with the 5.1 channel appellation. In home systems, the LFE (low-frequency effects) channel is blended with low frequencies from the five main channels, and the combined signal is handled by the subwoofer channel.

A current frustration for those who are experiencing the first waves of this technology, is that Dolby Digital is not always 5.1 channels. Many film and music DVD's are delivered in two-channel form encoded in a Dolby-Surround-like format. Some are old-fashioned stereo, and yet others are, wait for it, mono! This is all understandable, given the age of some of the artistic events. However, the poor – or non-existent - labeling of the packages, and the accompanying non-standardized on-screen menus, are enough to drive even the most enthusiastic of us to distraction. I have been in retail stores, where the staff thought that they were demonstrating digital discrete, when it was, in fact, familiar old Dolby Surround. I hasten to add that this confusion is not Dolby's fault, it is in the hands of the media distribution agents. Oh yes, at the time of writing this, Dolby Digital from satellite and cable is often (always?) in the two-channel configuration. All of this can only improve. Let us hope it does soon.



In Europe, the MPEG2 audio standard includes multichannel audio that can be either five or seven channels. In the seven-channel mode the additional channels are interpolated between the center and left and right front channels. For home applications it is difficult to imagine this becoming popular. A better use of the bandwidth might have been to add some truly rear channels.

Digital Theater Systems (DTS), and Sony's SDDS system have established themselves in the professional domain as the multichannel sound options for numerous films. On the consumer side, DTS encoded film soundtracks are available on laserdiscs and DVD's. DTS has also been promoting its system for music, and there are several multichannel audio CD releases in circulation. DTS also employs "lossy" data reduction, but it operates at a much higher data rate than Dolby Digital – using more bandwidth. The claims are that this yields superior sound quality, which indeed it should, but opinions vary as to whether the differences are consequential.

All of these discrete systems are really transparent transport media; none of them incorporates or is based on an underlying method for encoding and decoding spatial information. All of the matrix systems discussed up to now, put serious constraints on the creative process because of the cross-channel leakage, and steering artifacts. Discrete systems have no such limitations and, in fact, recording engineers have had to learn new techniques, and need new production tools, to recreate some of the illusions with which we have become familiar in the matrix systems. In short, we have entered a new realm of multichannel entertainment wherein what we hear will be almost entirely the result of individual creative artistry in the recording process and how this interacts with the particulars of the playback systems. Since there are no standards whatsoever, we can expect considerable variety in the results, including some examples of extremely bad taste. Be prepared.

Treated as multichannel transport media, however, these systems are potentially wonderful. They can be used to store audio data encoded in forms to entertain large audiences (like conventional film soundtracks), or audio data intended to reconstruct a 3-dimensional sound field (like the elaborate forms of Ambisonics), or for formats yet to be invented. They represent a freedom we have never before had.

All of these systems are scaleable, i.e. they can be designed to fit into different channel or storage capacities. There are two ways to achieve this, and both are used. Lossless data compression makes use of redundancy and signal variability to fit information into less "space" and then recover it, perfectly, during playback. This is the principal underlying MLP, the data reduction scheme that will be used in the Audio DVD, when and if it becomes a commercial reality.

Perceptual encoding achieves data reduction by taking advantage of both simultaneous and temporal masking in our hearing systems. It is well known that loud sounds prevent us from hearing weaker sounds. If we know the "rules" governing this phenomenon, we can simply eliminate or, at least more simply encode, those small sounds that are normally masked. Either way we can attempt to store the same *perceived* sound in less digital space. The more aggressive the data reduction, the more likely that listeners will be aware that the signal has been modified - something important has been edited out.

High-end paranoia would have it that all perceptually-encoded systems are fatally flawed, alluding to the discarded musical information. Well, it is only lost, if it could have been heard. Serious subjective evaluations by experienced and trained listeners have been involved with the optimization of these algorithms, to ensure that critical data are not deleted. These are in tests where listeners can repeat musical phrases and sounds as often as necessary for them to be certain of their opinions. Having participated in comparative listening tests of some of these systems, I can state categorically that the differences among the good systems at issue here are not "obvious". Even in some very aggressive data reduction systems, audible effects were quite infrequent, and limited to certain kinds of sounds only. And then, the effects were not always describable as better or worse, sometimes they could only be identified as being "not quite the same".

Naturally, it is possible to go too far, and in the most extreme examples of data reduction, things start sounding pretty bad. We can find examples of this in most of the sounds currently emerging from the internet. Needless to say, there is no reason to encumber our audio futures with systems that are annoying to listen to, or that lose otherwise audible musical information. However, I was frankly amazed at just how durable our auditory processes are, and concluded that perceptual encoding, *if it is done well and in moderation*, is not a fatal flaw - in fact it may not be detected at all.

In retrospect, one should not be totally surprised. After all, we have lived many years with LP's that performed "data expansion", adding unmusical information in the form of crosstalk, noise and distortions of many kinds. However, because of those very same masking phenomena that allow perceptual data reduction systems to work, the distortions were perceptually attenuated. So successful is this perceptual noise and distortion reduction, that good LP's played on good systems can still sound impressive.

Fortunately, in the digital domain, all things tend to become possible at lower prices and with higher speeds. With the end of this trend not yet in sight, it may be that the need for data reduction in critical applications will simply disappear.

WHERE WILL IT END?

Those who remember the quadraphonics debacle get a little queasy when they see what is going on now. Could this wonderful progression to digital discrete multichannel be stalled or stymied by too many options and a lack of agreement? The technical answer is no. The reason is that now we are operating in the digital domain and things are fundamentally different from the days of hard-wired analog algorithms. Nowadays, just as in a computer, one calls up a new program, and a different form of decoding can occur.

Playback devices read a code embedded in a program of any kind and, in response, configure itself to do the appropriate kind of decoding. Right now, only a few very expensive surround decoders offer forward compatibility – the ability to be upgraded to include new playback algorithms as they become available. However, just look at what has happened to computers, and think of these devices as special-purpose audio computers. With internet connections, the future will almost certainly include the ability to download music and movies, and any special "programs" necessary to play them. It is a very different world.

Digital discrete multichannel storage capability should not carry with it any restriction to, nor should it modify, the signals that are stored. In the multichannel matrix systems that was not the case - the encoding was part of the storage process. Now, it should be possible to envisage a system that could store some number of two-channel programs (stereo, binaural, or Dolby Surround, for example), or a four-channel version of Ambisonics and a two-channel program, or a 5.1, 7.1, or 10.2 channel discrete program, or . . .

Suffice it to say that, because technology is changing, it is now not necessary to establish hard universal standards. We could have several formats, each optimized for different applications ranging from uncompromised professional and high-end audio formats, to those that have been "scaled" in various ways to fit the cost and bandwidth limitations of portable, broadcast or network distribution media. In the short term, there will likely be some angst but, in the long term, these are technical problems that will find appropriate and affordable solutions. Place your bets now.

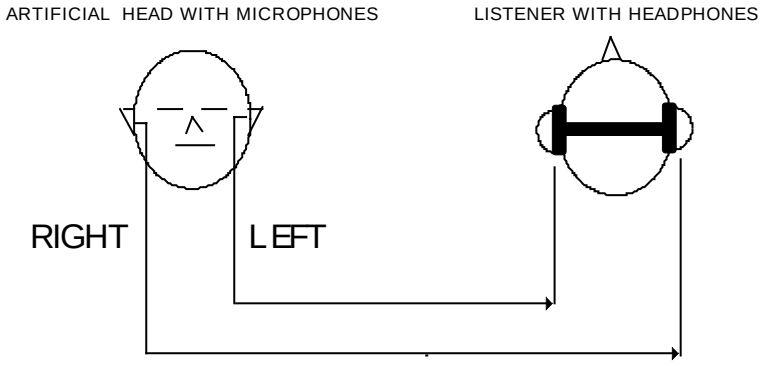
Part Two: Binaural Techniques

It seems that every decade or so binaural sound enjoys a revival. It was first demonstrated around the turn of the last century, and one can only imagine how bad it sounded then. Since that time, technology and knowledge have permitted ever better versions but, even today, it is not known or understood by most people. However, it is rapidly becoming widespread in the form of binaural 3-D audio in interactive computer games Many of us can play DVD's in our computers, and enjoying Dolby Digital multichannel sound reproduced through five phantom loudspeakers generated by sounds from only two real loudspeakers or headphones.

Picking up on the theme of Part One of this article, Ambisonics is a system that attempts to capture a 3-dimensional sound field and, in an idealized system, to immerse a listener in a facsimile of that sound field reconstructed by several loudspeakers. Binaural techniques attempt to capture the spatially-encoded sounds that enter the earcanals of an artificial "listener" at a live event, and then deliver those same sounds to the respective ears of real listeners, thereby reconstructing the *perception* of the original 3-dimensional sound field at different times and places.

A BRIEF HISTORY OF BINAURAL SOUND

Binaural means "two ears". When we listen through our two ears, we are hearing in "3-D" - perfect 3-D! All of the acoustical information needed for three dimensional auditory illusions is in the sounds arriving at our ears. Therefore, if we could encode sounds in the appropriate manner and reproduce them for each of our ears, we should be able to reproduce 3-D audio experiences.



TRADITIONAL BINAURAL "3-D"

It has long been acknowledged that, in theory, the most accurate recording technique is "binaural", in which the ears of an accurately modeled mannequin are fitted with microphones, the left and right ear signals are recorded and subsequently replayed through headphones to the ears of a listener. The Bell Labs study of auditory perspective concluded that in 1934 [2].

Ideally, the listener should experience an auditory illusion identical to that which would have occurred if he or she had replaced the mannequin at the original performance.

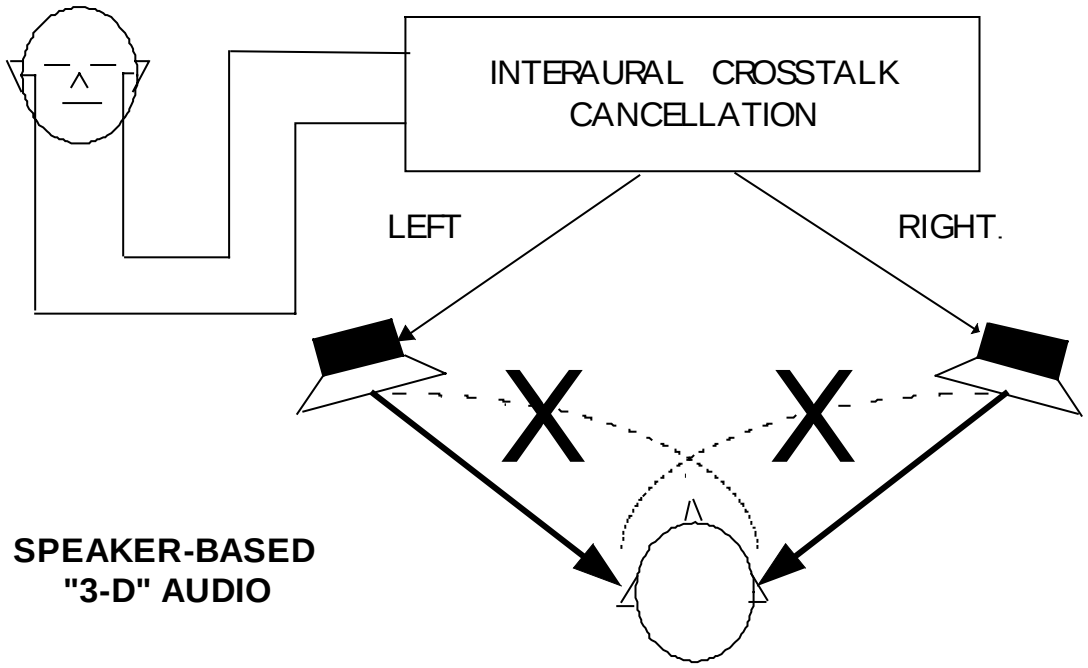
As it turns out, it works very well, but not perfectly. Listeners usually report a pleasantly spacious illusion, but sounds that should be perceived to be far out in front are instead localized inside or very close to the head, or even behind. Since most sounds of interest are outside, in front of us, within our field of vision, this is a serious problem.

Those of us who have been around audio for a while can remember numerous demonstrations of whispering in the left ear, and the right ear and noises like clipping hair at the back of the neck. Through headphones they sound very realistic. Distant sounds to the side and rear are also convincing. However, although some people report that voices and noises can be heard convincingly outside and in front, for most it simply doesn't work.

These perceptual errors have been attributed to a number of things: the lack of a visual confirmation for what is heard, the fact that the auditory illusion moves to follow head rotations, the fact that the mannequin's ears probably are not exact replicas of the listeners', headphone performance errors, and so on. In a static listening situation, the eye-ear-brain system is not fooled. Adding correlated visual cues and a dynamic head-position tracking system with the appropriate DSP corrections, is a great improvement, as has been demonstrated in the best of the virtual reality (VR) systems. Personalizing the system to match the listener's ears is another possibility. However, with all of that, it has to be said that this form of binaural reproduction is probably not yet a solution for the masses.

What is really needed is a delivery system in which the sound sources are convincingly external and in front. We need to be able to reproduce binaural signals through loudspeakers, but the problem is that the sound from each loudspeaker travels to both ears - there is "crosstalk" from the left loudspeaker into the right ear, and vice versa.

However, if we know where the loudspeakers are located, and where the listener is located relative to them, then it is possible to calculate or measure the unwanted crosstalk components. Then, in a signal processor upstream of the loudspeakers, the sounds can be processed so that, when they arrive at the ears, there is an acoustical "algebra" resulting in the left and right loudspeakers communicating independently with the respective ears.



The crosstalk-cancellation concept was patented by Atal and Schroeder at Bell Telephone Laboratories in 1966 [10], and used by them to study concert hall acoustics by using binaural recordings done in different venues, reproduced through loudspeakers in an anechoic chamber. It worked, but it required a massive computer, an anechoic chamber, and a listener locked into a sweet spot.

Versions of crosstalk-cancellation have appeared over the years, mainly as methods to expand the perceived soundstage from two-channel stereo. Of these, probably the best known were Carver's Sonic Holography [5] and Polk's SDA loudspeakers [11]. Lexicon's surround processors also have a binaural panorama mode which is a crosstalk canceller.

In the 1980's Duane Cooper and Jerald Bauck focused on the original problem of accurate binaural playback and developed a series of improvements that made speaker-based listening more practical and economical. These patented innovations resulted in a system that is simpler to implement than the Atal and Schroeder original, that is less demanding of the listening environment, that is more tolerant of head movements, and that degrades "gracefully" as the listener moves out of the sweet spot [12]. The Cooper-Bauck Transaural technology provides the basis for the Harman VMAx (Virtual Multi-Axis) system.

In the best systems, the sweet spot or, more accurately, the sweet region, is about the same as it has been in stereo for the past forty-odd years: long, tall and narrow. The difference is in the auditory reward. In stereo we get to hear the featured artist floating midway between the speakers. In speaker-based binaural, we can be transported to another three-dimensional world.

Obviously the system works best when the listener is in the predominately direct sound field of the loudspeakers. This means that close listening, as in a computer workstation, is likely to work well. At greater distances one must pay attention to reflected sounds, which can be done by controlling the directivity of the loudspeakers or the absorption of the reflecting surfaces in the room. Best performance will always be

observed when the listening geometry matches that for which the crosstalk-cancellation filters have been designed.

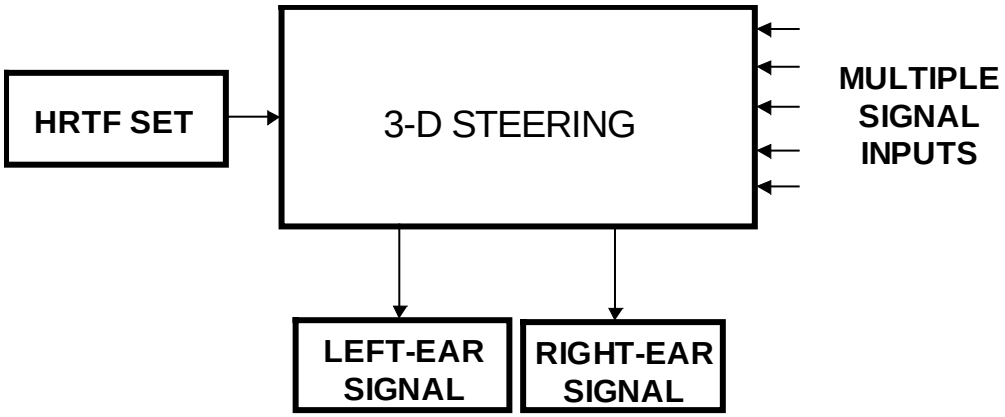
Since the sounds come from loudspeakers, head movements simply confirm that the sounds originate outside and in front of the listener. In headphone reproduction of binaural signals, it is difficult to create convincing illusions of sounds originating in front of the listener. Here the difficult area is behind the listener; on balance, this is the preferable compromise. However, in practice, when the images are in motion, and if there are visual cues that correlate with the sound movements, most people drift into a susceptible frame of mind in which even these front-back uncertainties disappear.

That there should be these front-back reversal problems in headphone and loudspeaker reproduction of binaural programs is not surprising. The problem is the location of our ears, and the front-back symmetry that exists. Auditory cues alone are not enough to give us a completely reliable front or back designation. In tests of natural localization capabilities, front-back reversals are frequently observed. In the course of our everyday lives we rely on head movements and visual cues to keep things straight. Removing or altering those normally-reliable cues makes things go perceptually wrong. Since much of our natural auditory localization relies on “plausibility”, we must conclude that the perception of a sound source outside and in front is less plausible in headphones than it is through crosstalk-canceled loudspeakers. The reverse is apparently true for sound sources behind. In-head localization must therefore occur when nothing else appears to be plausible. [13][14][15]

BINAURAL STEERING - SIMULATING THE ARTIFICIAL HEAD

This is all very well but, so long as the recording process requires an artificial head, 3-D audio will not be popular - the recording industry is committed to multi-track, multi-mike recording methods. With DSP it is now possible, in real time, to electronically synthesize the left- and right-ear signals appropriate for any direction.

To achieve this we must know, for every direction we wish to synthesize, how the sound is modified on its way to the ears. This is learned by positioning a known sound source at various points in space around a head, and measuring the “head-related transfer functions” (HRTF's) to each of the ears. The HRTF’s are measured as impulse responses (amplitude versus time), or as the Fourier equivalent: amplitude and phase versus frequency. With this information stored away, the binaural directional synthesizer can modify any single-channel signal into the left- and right-ear signals appropriate to the chosen direction.



BINAURAL SIGNALS TO HEADPHONES OR
CROSSTALK-CANCELLED LOUDSPEAKERS

BINAURAL DIRECTIONAL SYNTHESIS

This process has been widely used since about 1980, when affordable computers of sufficient speed and power were available. Among the best known efforts of this kind is the collaboration among Elizabeth Wenzel at the NASA-Ames Research Center, Fred Wightman and Doris Kistler at the University of Wisconsin, and Scott Foster at Crystal River Engineering who focused on headphone reproduction for the military and virtual-reality applications of the technology. Another pioneering effort was that of Bo Gehring of Focal Point 3D Audio. Both of these systems, used with head-position tracking, provided quite convincing 360-degree localizations. Several more processors now exist that can do this and, thanks to the MIT Media Lab, a set of HRTF's is now available on the web. Durand Begault provides a detailed document of this subject in his book [16].

So, there it is. We have two ways to create binaural signals, dummy head microphones and electronic synthesis, and two ways to reproduce them, headphones and crosstalk-canceled loudspeakers.

THE NATURE OF THE "SWEET SPOT"

In <u>any</u> system involving binaural image steering through loudspeakers, there is a sweet spot - a location at which the 3-D sound "picture" is most sharply in focus. Systems claiming to have a large sweet spot do so at the expense of localization precision: "fuzzy" sweet spot = "fuzzy" localization. For some kinds of entertainment that may be sufficient, for others it is not.

In general, the sweet spot is long, tall and very narrow. Listeners will find that there is considerable tolerance to front-back movement, and to vertical movement (determined mainly by the vertical directivity of the loudspeakers), but little tolerance to sideways movement. Systems will differ in their tolerance of head rotation, and to movement from side to side, off the axis of symmetry. Ideally, angular variations of ±20 degrees or more should not dramatically change in the illusion. Small movements from side to side should cause the soundstage to distort in an "elastic" fashion. Larger movements away from the axis of symmetry should result in a smooth degradation from a

3-D illusion to one with reduced dimensions. At no time should the listener be aware of a “pulling” sensation or obvious “phasiness” in normal small movements of the head.

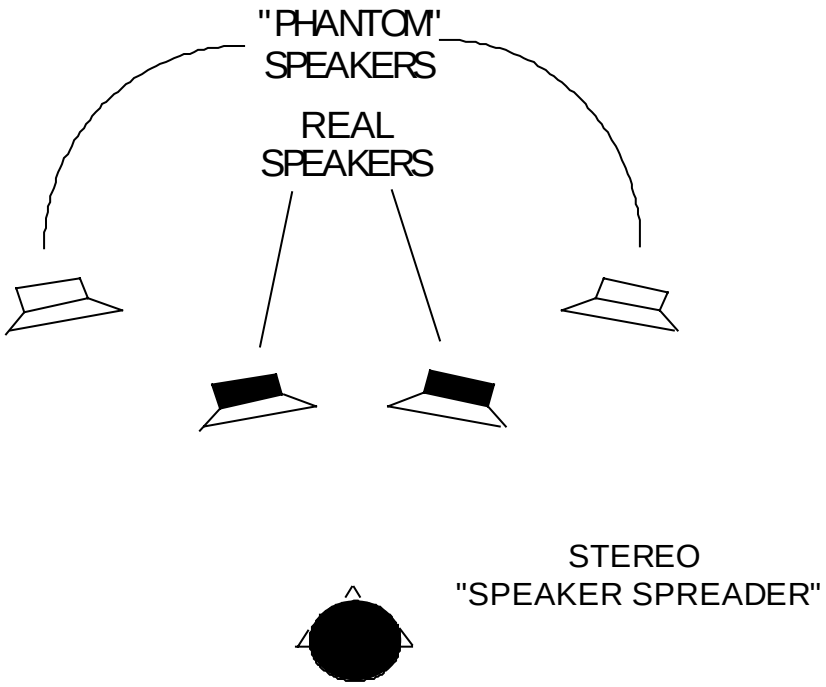
A sweet spot, or more accurately a "sweet region", is not new to us - there has always been one for conventional stereo. That most stereo listeners ignore the “sweet spot” is a measure of the marginal reward for sitting there. In crosstalk-canceled loudspeaker 3-D audio, the rewards are enormous if it is well done.

Not everyone wants to put up with the restrictions of a sweet spot. Not everyone needs to have the luxury of precise image localization. For those whose needs include both of these, there is a solution. By simplifying the process, corrupting it some would say, it is possible to create systems that generate a gratifying sense of envelopment and space, but that cannot deliver the perceptions of images precisely located where they should be. When this is done, one finds that the acceptable listening area is quite large.

So, binaural systems, and approximations to them, are scaleable. If one accepts the restriction of a sweet spot, then the directional and spatial illusions can be remarkably precise and realistic. As the sweet spot is enlarged, the localizations become less well defined, but a pleasant spaciousness can be preserved.

THE REPERTOIRE OF SPEAKER-BASED BINAURAL AUDITORY EFFECTS

1. SPEAKER SPREADER. In those circumstances where the loudspeakers are too close together to yield a realistic stereo soundstage, this kind of processing makes it possible to replace the real loudspeakers with phantom loudspeakers having an increased angular separation. Done well, the effect is so convincing that little or no sound is perceived to come from the real loudspeakers. The angular separation can be varied according to the listener's taste. An additional perceptual advantage of listening through phantom loudspeakers is that the impression of depth in the stereo soundstage can be enormous, since it is not tied to the distance of the real loudspeakers, but dictated more by the distance cues in the recording itself.



2. CROSSTALK-CANCELLATION. Used alone, acoustical crosstalk cancellation permits the left loudspeaker to communicate with the listener's left ear, and the right loudspeaker, with the listener's right ear. In this sense the delivery system is not unlike headphones, but there is an important advantage - the sounds are perceived to be external and in front of the listener.

Binaural (e.g. dummy head) recordings can be played directly through this system and the result can be a remarkably convincing sense of being in a different three-dimensional space. Not all binaural recordings are equally good. Differences among the artificial-head microphones, post processing and positioning *vis a vis* the performers create clear preferences among the few existing commercial examples. The good ones sound stunningly realistic.

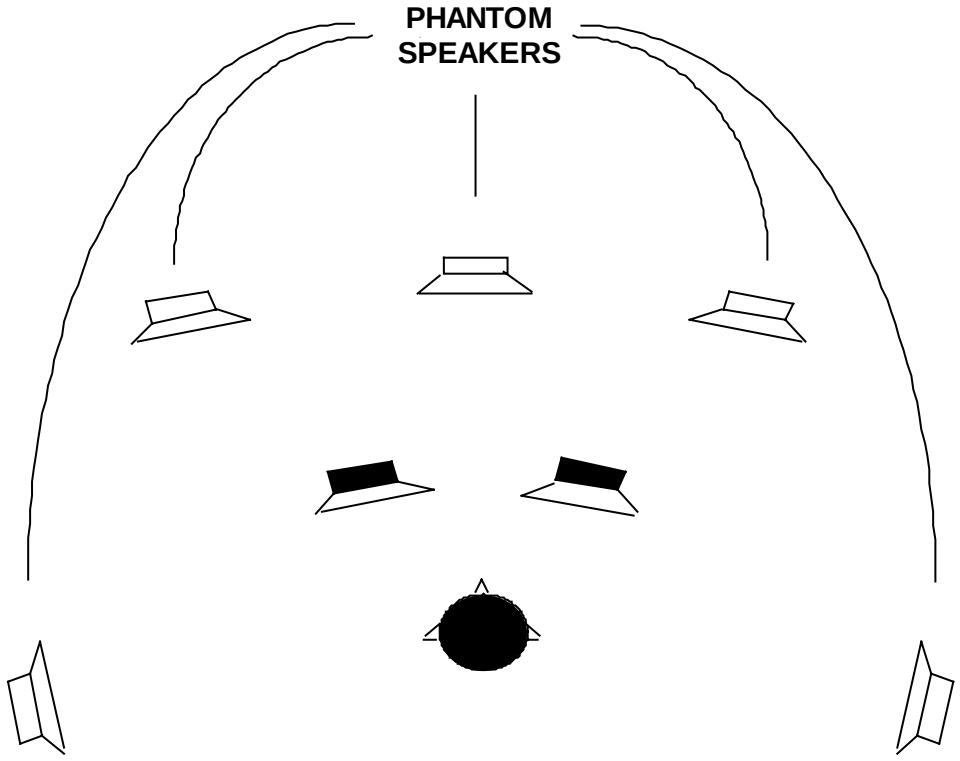
Conventional stereophonic recordings also can be played through the system. The results will vary, depending upon the microphone techniques and signal processing employed in the making of the recording. My experience has been that a high percentage of recordings, especially popular recordings, are enhanced in very interesting ways, sometimes expanding to fill the entire front hemisphere. Some recordings exhibit very dramatic and pleasant three-dimensional illusions, much more engaging than is possible with conventional stereo reproduction. Nowadays, a few recordings are preprocessed with "3-D" effects. These may or may not be compatible because of double processing.

3. “PHANTOM” HOME THEATER. In this scheme, the five channel outputs from a surround processor are connected to a five-axis binaural steering device. The channels are then steered to the appropriate locations in space around a listener: say, at 30 degrees left and right and zero degrees for the front channels, and 90 degrees left and right for the surround channels.

This done, there are two options for playback: headphones and crosstalk-canceled loudspeakers. With signal conditioning to account for peculiarities of the headphones,

one can then experience a multichannel simulation through headphones. All of the preceding provisos for headphone listening apply, principally, the likelihood that sounds that should be outside in front may be perceived to be very close, or inside the head. The experience may be a compromised one but, on balance, it is likely to be better than the stereo version.

If we take the next step and create a crosstalk-canceled loudspeaker version, then the listener can have a multichannel experience, from any source, through just a single pair of loudspeakers. Here, the difficulty is in getting convincing sounds behind the listener something, fortunately, that is rare, even in movies.



A "PHANTOM" HOME THEATER

With the digital discrete multichannel formats like Dolby Digital (AC-3), all five channels are equal - it is no longer possible to get away with lesser performance in any of the channels. Each of the five channels can be used for a convincing directional effect. Speaker-based binaural synthesis can solve the problem by creating a "phantom" multichannel system in which all of the channels are identical in sound quality, and each one can be addressed independently. The loudspeaker/amplifier budget can be concentrated in two good loudspeakers, rather than distributed among five lesser products.

When implemented as described here, these synthesized multichannel techniques are not gimmicks; they are not adaptations of or modifications to two-channel stereo, Dolby ProLogic, Dolby Digital, or any other multichannel input that is selected. That

decoding is done *before* the image steering, using processors approved by the respective systems. All that is done here is to synthesize phantom loudspeakers to replace real ones.

Sadly, there are systems in the marketplace, masquerading as phantom home theater, that cannot create the five independent phantom sources. Such systems are instantly revealed by playing test signals through the system, in which each channel is exercised separately.

4. GAMES AND INTERACTIVE ENTERTAINMENT. The ability to binaurally steer specific sounds to specific locations in space can provide interactive games with a truly impressive enhancement. Good guys and bad guys can now be audibly tracked and chased. Full interactivity ensures that, as the player alters the visual perspective, the sounds in space will track correspondingly to the correct locations. This has been demonstrated over headphones in various helmet-based virtual-reality games that have used a head-position-tracking device to provide spatial interactivity. Speaker-based binaural provides a parallel experience in computer workstations. Harman's VMAx algorithm is part of Microsoft DirectX 7.0 – 3-D audio for games.

CHALLENGING CONVENTIONAL WISDOM

Obviously, the quality of our auditory perceptions is dictated by the integrity of the sounds arriving at our ears. Conventional stereo and multichannel sound systems use several loudspeakers to locate sounds in space and to energize the reflective sound field in rooms. The room is very much a part of the sound reproduction system and, as such, represents a substantial uncontrolled variable in the critical final phase of the audio delivery system. The physical locating of loudspeakers and the acoustical treatment of listening rooms has become part of the personal art of audiophile stereo systems. In the end, they are all different. Close-up listening to small loudspeakers (often called near-field listening) has become increasingly popular in professional recording as a means of reducing the influence of variable room characteristics prevailing at different studios. It gives the engineers a better *absolute* reference for sound quality than the traditional large monitors which are subject to the variability of rooms. It is a trend which also fits nicely with the audio/video workstations which are becoming increasingly common in production and post-production tasks in film, television and, of course, multimedia and computer games.

In multichannel surround sound systems, a persistent problem is the mismatch in the timbral signatures of the various loudspeakers. Some of this may be caused by real differences between the loudspeakers but, even if they are identical devices, there will be differences attributable to different positions in the room. With crosstalk-canceled loudspeaker systems, all of the sound comes from two loudspeakers so, by definition, all phantom loudspeakers in virtual home theater systems differ in timbre only by the differences in HRTF's associated with their different locations. In other words, it is like listening to five perfectly-matched loudspeakers in a perfect room.

A further non-trivial attribute of this form of listening is the remarkable sense of distance and depth. With the crosstalk canceled, the listener has no information by which to judge the distance of the real sources of sound - the loudspeakers. Impressions of distance, then are derived from cues that are in the recordings. It is captivating to see

loudspeakers, but not to “hear” them and, instead, to be aware only of sounds occupying positions or areas in a perceptual space that extends far beyond the walls of the room.

So, let us adjust our mindsets slightly. Let us not think of small loudspeakers in close listening situations as poor substitutes for the traditional professional and hi-fi products. Let us view them as legitimate alternatives that, in some important ways, have the potential of being even better.

TRUE SUCCESS IS IN THE ATTENTION TO DETAILS

Convincingly good 3-D audio requires that there is a well-defined acoustical link from each of the loudspeakers to each of the listener's ears. Notions that any pair of loudspeakers stuffed into any convenient location will work, are naive. Disappointments will abound. Fortunately, with attention to the right details, it is not difficult to be very successful.

First, the loudspeakers must be closely matched in performance, and the electrical signal paths balanced.

Second, the listeners must be in a predominantly direct sound field. Sounds reflected from nearby objects - walls, tables, the work surface, monitor, computer, printer, keyboard, and the like - corrupt the illusion. Controlling the directivity of the loudspeakers to avoid reflections is a further benefit. In the latter case horizontal as well as vertical directivity control is advantageous.

Third, the left and right acoustical signal paths must be the same length - the listener must be close to the axis of symmetry of the loudspeakers, equidistant from both.

The computer workstation environment is one that lends itself, naturally, to a solution to these concerns. However, with the appropriate attention to the design of the loudspeakers and listening environment it is possible to provide excellent experiences at much larger distances.

In all instances, it is assumed that the listener has normal hearing in both ears. If this is not true, the system breaks down. It is also assumed that the listener’s head and ears have acoustical properties that are similar to those embodied in the HRTF’s used for the binaural steering and crosstalk cancellation. While there are powerful features that we have in common, there are certain to be instances in which different people hear slightly different things from these systems. No one listener, therefore, can be the arbiter of absolute quality. Experience shows, however, that it is possible to design a system that can provide remarkably convincing performances for most listeners.

CONCLUSION

So, there it is, my version of where we are, and where we might be going. And we have some answers to the question in the title: how many channels do we need to believe that we are “there”? It can be several – 5 or 7 or more – or it can be just two, as in binaural, not stereo, techniques. I have tried to make clear the distinction between those systems that are “social” in that they allow for multiple listeners, and those that are

designed to satisfy only a single listener. There are clearly times and places for both kinds. Technologies now exist that can give us a choice, and there are undoubtedly more to come.

I find the whole thing very exciting because, at last, we have broken the two-channel stereo stalemate. We knew long ago that there were better ways. As a stopgap (a forty-five year one!) stereo has been very enjoyable. It is time to move on, though. Are you ready?

REFERENCES

[1] British Patent No. 394 325 "Improvements in and relating to Sound-transmission, Sound-recording and Sound-reproducing systems" granted to Alan Blumlein and EMI, 1933. This is reprinted in J. Audio Eng. Soc., vol.6, pp. 91-98, 130, April 1958.

[2] Steinberg, J.C. and Snow, W.B. "Auditory Perspective - Physical Factors", Electrical Engineering, vol. 53, PP. 12-17, Jan. 1934.

[3] Snow, W. B., "Basic Principles of Stereophonic Sound", IRE Transactions - Audio, vol. AU-3, PP. 42-53, March-April 1955.

[4] Hafler, David "A New Quadraphonic System", Audio, July 1970.

[5] Carver, Robert "Sonic Holography", Audio, vol.66, March 1982.

[6] Nakayama, T., Miura, T., Kosaka, O., Okamoto, M., and Shiga, T. "Subjective Assessment of Multichannel Reproduction", J. Audio Eng. Soc., vol.19, pp. 744-751, Oct. 1971.

[7] Cooper, D.H. and Shiga, T., "Discrete-Matrix Multichannel Stereo", J. Audio Eng. Soc., vol 20, pp/346-360, June 1972.

[8] Gerzon, Michael, "Ambisonics in Multichannel Broadcasting and Video", Presented at the 74th Convention A.E.S, Preprint 2034, Oct. 1983.

[9] Web sites with comprehensive bibliographies and other useful information on Ambisonics and related subjects: www.users.dircon.co.uk/~wareing/3daudio.htm

[10] U.S. Patent No. 3,236,949 "Apparent Sound Source Translator" granted to B.S. Atal and M.R. Schroeder, assignors to Bell Telephone Laboratories, Inc., 1966 (Filed 1962).

[11] Polk, Matthew, "Polk's SDA Speakers, Designed in Stereo", Audio, vol.68, June 1984.

[12] Cooper, Duane and Bauck, Jerald, "Prospects for Transaural Recording", J. Audio Eng. Soc., vol.37, pp.3-19, Jan.-Feb. 1989.

[13] F.E. Toole, "The Acoustics and Psychoacoustics of Headphones", 2nd International Conference, Audio Eng. Soc., preprint C1006 (1984 May).

[14] F.E. Toole, "Binaural Record/Reproduction Systems and Their Use in Psychoacoustic Investigations", 91st Convention, Audio Eng, Soc., preprint no. 3179. (1991 October).

[15] F.E. Toole, "In-Head Localization of Acoustic Images", J. Acoust. Soc. Amer., Vol. 48, pp. 943-949, 1970

[16] Begault, Durand, “3-D Sound for Virtual Reality and Multimedia”, AP Professional, 1994.

March 5, 2000

This paper is an updated version of a two-part article entitled “The Future of Stereo” which appeared in the May and June 1997 issues of Audio magazine.