Direction and Space – the Final Frontiers

Multichannel and binaural audio systems for movies, music and games

by Floyd E. Toole, Ph.D.
Vice President Acoustical Engineering, Harman International Industries, Inc.
8500 Balboa Boulevard, Northridge, CA 91329   www.harman.com

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 in cinemas, homes and cars. Is this the solution that we have been searching for?

Here we review some of the past, examine the present, and speculate about the future of multichannel audio and what it can do for us all.

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 1930s 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 most elementary 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."

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 suf-

1. Monophonic means reproduction through a single channel and loudspeaker. Monaural means, literally, one ear, but is frequently used interchangeably with monophonic. We listen binaurally, through two ears, however many channels are active.

fered 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 AUDIO – A HUMBLE BEGINNING

The arrival of two-channel stereo in the '50s 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".

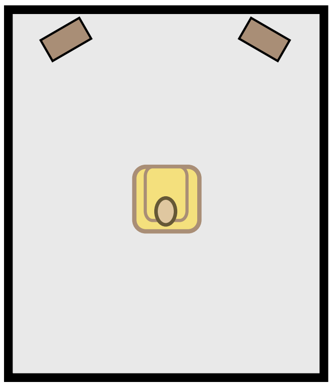

Figure 1. Multichannel audio: the first attempt. Two-channel stereo as we have known it for over 50 years. It is optimized for a single listener in a symmetrical "sweet spot" – a restriction that is routinely violated.

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. Greybeard audiophiles will remember the Hafler system [4] sold by Dynaco, decades ago, an early attempt at ambience extraction using an additional loudspeaker. And then there was Carver's Sonic Hologram a simplified binaural crosstalk canceller [5] that expanded the soundstage beyond the left and right loudspeakers. Nowadays, hosts of spatial processors are available to add 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.

2 + 2 = Quad

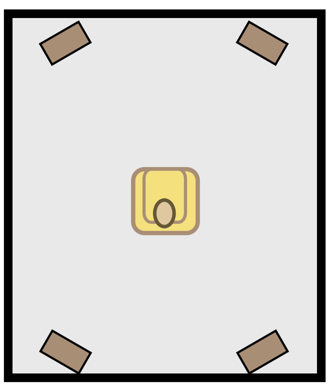

Figure 2. Multichannel audio: the second attempt. Quadraphonics. The "sweet spot" is now restricted in two dimensions.

In the '70s 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 LPs 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 LPs. 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 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 get some relief from the restrictions of 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 clever technical innovation in quadraphonics would live on.

MULTICHANNEL AUDIO - CINEMA MAKES IT WORK

Figure 3. The Academy of Motion Picture Arts and Sciences, Samuel Goldwyn Theater, showing five front channels of JBL loudspeakers and eight 18-inch subwoofers for low-frequency effects (LFE). All of this, of course, is hidden behind a perforated screen. Twenty-four surround loudspeakers complete the installation.

Failure in one market application 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.

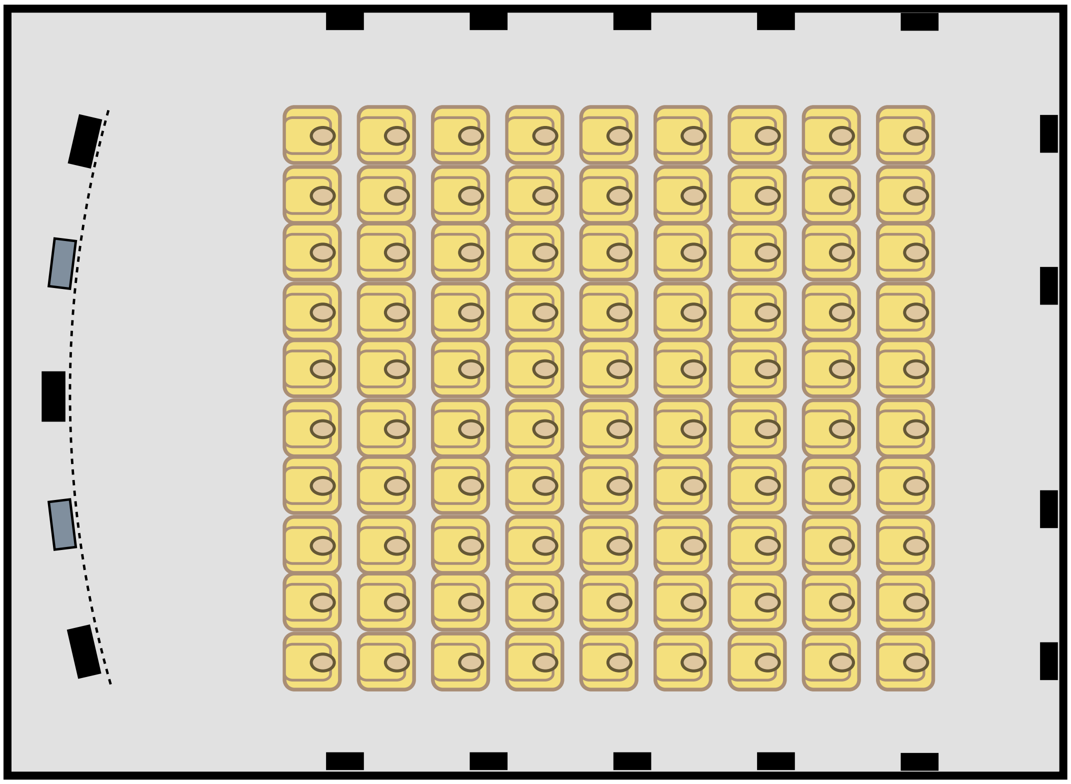

Figure 4. A multichannel cinema system, showing the standard left, center and right screen-channel loudspeakers with additional center-left and center-right units for SDDS sound tracks. The array of surround loudspeakers, down the sides and across the back, is driven by the same signal when the source is Dolby Stereo, and is split into left and right groups for 5.1 channel sound tracks. The rear surrounds can be separately driven for those films with 6.1 channels.

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, gunshots, ricochets and the like. At times the audience can be:

- ❐ enveloped in sound as in a football game – multidirectional sounds,
- ❐ transported to a giant reverberant cave or gymnasium – lots of reverberation,
- ❐ in the confines of a car engaged in a dramatic chase – multidirectional sound, small room reverberation,
- ❐ treated to an intimately whispered conversation – no surround or ambiance.

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 are 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.

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.

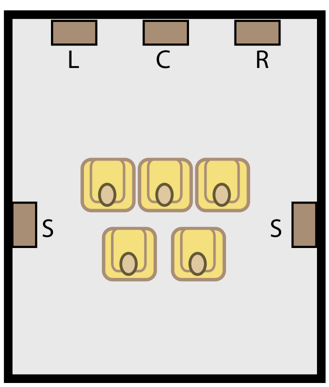

Figure 5. Multichannel audio: the third attempt. Dolby ProLogic, a four-channel active matrix system. It is social!

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 were better, but even they 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

In a natural succession to their THX program for certifying cinema sound systems, Lucasfilm established a licensing scheme for features intended to help in reproducing a cinema experience in home theater systems. Home THX, as it was called, added features to a Dolby ProLogic processor and to the loudspeakers used in home theater systems. It also set some minimum performance standards for the electronics and loudspeakers. At a time when the market was inundated with "cheap and cheerful" add-on center and surround speakers and amplifiers, THX made a clear statement that all channels had to meet a similar standard. This amalgam of existing and novel features was a useful stimulus for consumer home theater in the early days. The THX embellishments in the original specification were:

1. High- and low-pass filters to help achieve a proper crossover between the subwoofer and satellite loudspeakers. This was a serious omission in early surround-sound systems, leading to lots of bad bass. Now, with even more flexibility, it is common in processors and receivers. In-room measurement and equalization is what is really needed.

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. As I recall, it was first introduced in the Shure HTS systems. Now that surround channels are discrete, the current adaptive version is activated only when a monophonic surround signal is detected, as in Dolby Surround sound tracks.

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 from random incidences will have timbres that differ from sounds arriving from the front because of the head and external ears – see HRTF explanation, p. 16. It is nature at work, and it needs no correction.

4. "Re-equalization" of the sound track to compensate for excessive treble that is sometimes built into film tracks to compensate for sound systems in large cinemas. THX used a single correction curve but it needed to be adjustable since sound tracks vary in treble balance. Now, movie and music concert DVD's are often mastered in circumstances resembling home theaters, so the problem goes away. It is a tone control, and 'off' is a useful setting.

In loudspeakers, THX required some control of the vertical dispersion from the L, C, R (front) units. To keep costs in line, the simple driver configurations that were sometimes used resulted in off-axis performance that no self-respecting loudspeaker engineer would be proud of. The requirement has been relaxed in recent years. Directivity control can be useful if it is done well, as it can be with horns or complex arrays of drivers, especially when combined with digital processing.

The recommended bi-directional "dipole" surround loudspeakers increased the proportion of reflected sound that is generated by the surround speakers, making it appear that there are more than two of them. Success is highly dependent on the specific acoustics of the listening room, since most of the sound is reflected. This was a persuasive argument in the days of Dolby ProLogic, with its single surround channel. Now, discrete left and right surround channels allow for acoustical and electronic decorrelation between the left and right surrounds in the program itself. The additional surround channels in 6.1 and 7.1 systems are more reasons to consider alternatives.

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 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. 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 Gerzon 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 problem. However, there are other considerations. 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.

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.

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. Binaural sound anyone? See Part 2.

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.

Perhaps the greatest conflict with THX concepts arises when a customer wishes to reproduce multichannel music, since this is recorded with the expectation that playback will be through five similar loudspeakers. Loudspeaker manufacturers have addressed the situation by selling surround loudspeakers with other directional patterns, notably a bipole configuration (bidirectional in-phase), and with switch-selectable directivity – monopole, bipole, dipole.

Figure 6. An Infinity Beta ES250 wall-mounted surround loudspeaker, with a switch to select the directional mode: monopole, bipole or dipole. The customer can use the mode appropriate to the application, the room, or personal preference. The two active faces can also be driven independently, for special situations.

The situation is not static. Perhaps in a reflection of the interests of the new ownership, THX has been pursuing other product areas to certify, including computer and car audio, video displays, wire, DVD's, games, etc. The move to digital discrete multichannel delivery systems has substantially diminished the need for some of the original THX features. On the loudspeaker side, a good loudspeaker is a good loudspeaker, whether it is for movies or music.

THX certification is a licensing business, and certification of components is a manufacturer's option; it costs money. From the consumer's perspective, it should be regarded as an assurance of an acceptable level of performance in a product. However, many components that are not THX approved also perform to the same or even higher standards of performance. It is a marketing decision.

MATRIX MANIPULATIONS

Lexicon MC-12

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 play TV sound or 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 emphasis on “home theater” applications of multichannel audio systems has turned off 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.

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.

LOGIC7

The Logic 7 Algorithm

Dr. David Griesinger, at Lexicon, is probably best known in professional audio as the author of reverberation algorithms used in recording studios worldwide. 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

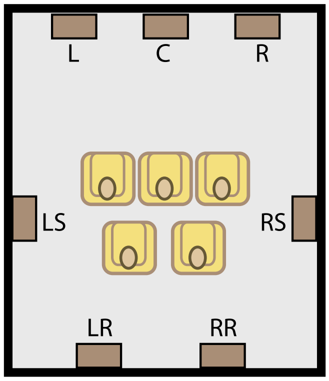

Figure 6. Multichannel audio: the fourth attempt. Logic 7 provides separation between left and right side surrounds and adds rear surrounds.

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.

He developed a fully digital implementation of Dolby ProLogic for incorporation into Lexicon's first consumer product – a surround processor. Eager to use the potential of an active matrix decoder for converting stereo music into a surround experience, he developed his own approach to decoding. The result is a suite of film and music playback algorithms embodied in Lexicon digital surround processors, and in other Harman International products from Harman/Kardon, JBL, and Mark Levinson. Called Logic 7, since it provides for two additional channels and loudspeakers behind the listener it can also be configured for 5.1 channels if required. All channels are full bandwidth.

Listeners 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. The more elaborate products allow users to adjust several parameters within the decoder, to cater to characteristics of different loudspeakers and playback environments, and to different listener preferences. Since stereo recordings were not created with this form of playback in mind, some of these adjustments will be program dependent.

Logic 7 also accepts 5.1-channel digital discrete signals from Dolby Digital and DTS sources and generates rear channel signals from them. The effect of adding the rear channels really has little to do with the impressions of sound coming from behind, although this occasionally happens. Instead, the dominant impression is of superior spaciousness, and an expanded seating area for quality listening.

Given Harman's active involvement in automotive OEM audio/navigation/infotainment systems, it is only logical that Logic 7 should evolve into versions suitable for automotive audio. We will talk more about this a little later.

While it is tempting to look for cause and effect, it is a fact that recently Dolby has introduced a system having the basic functions of Logic 7, called ProLogic IIx. DTS also has a version, called Neo:6. Clearly, this kind of processing has caught on.

DIGITAL AND DISCRETE MULTICHANNEL AUDIO

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 experience 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.

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 subwoofer channel handles the combined signal.

… even the most fastidious super-humans, dogs and aliens will be pleased.

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 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.

Adding to the confusion for consumers, are the competing and incompatible Audio DVD and SACD audio-only formats. Enthusiastic 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. However, even that is finite, and one wonders whether the effort would be better used in encoding more channels. A close listen will reveal that 5.1, 6.1 or 7.1 channels are inadequate to produce seamless directional and spatial sensations. Center left and right channels, as shown in Figure 4 would be nice, allowing us to expand the front soundstage, and some height information would bring us closer to reality. It is easy to imagine 10 or more channels. For customers who do not have all of the playback equipment, such systems can be designed to gracefully degrade to fewer channels. Something to think about.

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 experience considerable variety in the results, including some examples of extremely bad taste. Sound familiar?

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 scalable, 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.

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. To ensure that critical data are not deleted, serious subjective evaluations by experienced and trained listeners have been involved with the optimization of these algorithms. In these tests, 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, as some audio writers would have us believe. 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 hear this in some of the low bitrate 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. Perceptual encoding, *if it is done well and in moderation*, may not be detected at all.

In retrospect, one should not be totally surprised. After all, we have lived many years with LPs 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 LPs 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 audio data reduction in critical applications will simply disappear.

MULTICHANNEL AUDIO IN CARS

Traditional car audio is based on two channels – stereo. We play stereo program material in cars, but we do not hear stereo. Think about where the stereo seat is. Instead we manipulate interchannel amplitude and phase, trying to skew the soundstage to put the featured artist or radio announcer somewhere near the middle of the dashboard. It kind of works, but it is not what one hears at home. If we add the effects of package-shelf speakers in the rear, we end up with something totally different – car audio. That is what we all have been living with in our cars. Would it not be nice to hear in our cars something that resembles what we hear at home? We can: multichannel audio.

However, it is not altogether straightforward. Just think of what would happen if you shrank the listening room to the size of a large closet. It would sound very strange when you can reach out and touch several of the loudspeakers. The room is the final audio component, and recordings are made on the assumption that there will be reflected and reverberated sounds to augment those emerging from the loudspeakers. A car interior is a very intimate environment compared to any normal listening space. It is also acoustically 'dead'.

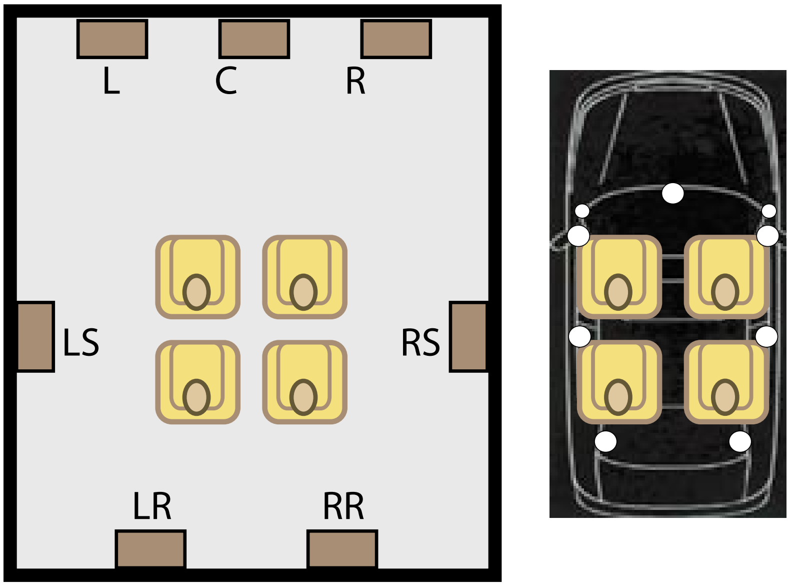

Figure 7. A comparison of a seven channel audio system installed in a home and in a car. The images are scaled relative to the four listeners.

What happens is that those digital discrete recordings that sounded so spectacular from any of the seats at home, now take on spatial distortions that are strongly associated with the listening location. Because of the proximity of loudspeakers – different ones for different seats – a system that has been balanced for the driver, for example, will sound very different in other seats. Back seats typically suffer most, and none of the seats will have the spatial perspective that existed in the listening room at home. There is no doubt about the "gee whiz" aspect to the listening experience compared to conventional car audio, but it still needs work.

Having had experience with active matrix decoded multichannel audio in cars (Logic 7), it becomes clear that the interchannel leakage (crosstalk) characteristic of those systems, is actually a good thing in a car. It softens the localizations, so that a saxophone isn't just screaming from a single loudspeaker at arm's length, it is shared with others nearby. What this implies is that some "repurposing" of discrete multichannel programs is required to make them more pleasant in car interiors and, especially, to help preserve some of the front soundstage illusion for those in the back seats. It will be interesting to see how this develops.

WHERE WILL IT END?

Stereo survived as an intact delivery format for over 50 years because it was an analogue, hardware-based, system. In the digital domain, change is relatively simple. In fact, we already have several multichannel formats, each optimized for different applications ranging from uncompromised professional and high-end audio, to those that have been scaled in various ways to fit the cost and bandwidth limitations of portable, broadcast or network distribution media. Even in inexpensive receivers, playback of these sources is automatic and transparent to the user. Systems are commonly configurable for 2 to 7 channels with or without subwoofers, thus catering to different budget and space constraints.

With internet connections, we can download all formats of music and movies, and any special decoding algorithms necessary to play them. Hard-drive audio and video PVRs and servers are widely available. In the short term, however, we still have to deal with "hard" delivery media: CD, DVD, DVD Audio, SACD, and several versions of HD video storage media. Compatibility among these is already a problem. In the long term, however, these technical problems will find appropriate and affordable solutions. Place your bets now.

Part Two: Binaural Techniques

CLASSIC BINAURAL AUDIO

"Any sufficiently advanced technology is indistinguishable from magic."

Arthur C. Clarke

When one hears truly well-done binaural reproduction, one experiences a justification for Arthur C. Clark's statement. Stereo is fun, *good* binaural truly is a magical experience. Sadly, good binaural audio is hard to find.

The Bell Labs study of auditory perspective pointed out the intrinsic excellence of binaural audio in 1934 [2]. Records indicate that it was first demonstrated around the turn of the last century, with what must have been seriously compromised sound quality. Since that time, technology and knowledge have permitted ever better implementations but, even today, it is not known or understood by most people.

Binaural means "two ears". When we listen through our two ears, normally, 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, see Figure 8.

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 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, above, 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 eerily realistic. Distant sounds to the side and rear are also convincing. In those demonstrations, one rarely heard anything that should be clearly in front, and especially at a distance in front. For most people, frontal sounds tend to be too close.

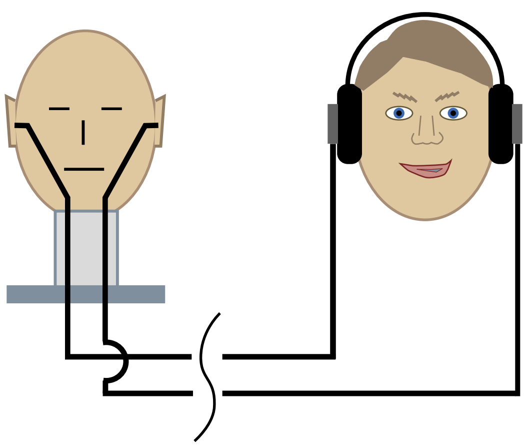

Figure 8. A basic binaural record / reproduction system. A dummy head, with anatomically-correct external ears, is equipped with tiny microphones. The recorded signals eventually make their way to headphones which reproduce the sounds for the appropriate ears. The listener should hear what the dummy head "heard".

These perceptual errors have been attributed to a number of things:

☐ the lack of a visual confirmation for what is heard. The eyes are known to dominate in many sound localization situations, e.g in movies most of the sound comes from the center channel and we generally don't notice,

☐ the fact that the auditory illusion moves to follow head rotations. In natural hearing this does not happen; head movements are known to be critical in resolving front-back directional ambiguities,

☐ the fact that the mannequin's ears are not exact replicas of the listener's external ears. This means that the detailed acoustical processing done by the external ear (pinna) of the mannequin will not be matched to the expectations of the listener. If head-related transfer functions (HRTFs) are really important maybe we need to have personalized recordings[2],

❑ headphone performance errors. The acoustical coupling between typical supra-aural (on the ear) and circumaural (around the ear) headphones and the eardrum is known to be quite variable, depending on the specific design [11]. Therefore, it is not possible to predict with accuracy what will be heard. This can be partially corrected for in some circumaural designs, or almost eliminated with in-ear headphones [12].

A lot has been written about the importance of the HRTFs and the need to personalize them in order to avoid front-back confusions and other localization errors. However, it seems to me that a couple of simple experiments settle the issue.

1. In an anechoic space it is possible to experience both in-head localizations and front-back reversals while listening with one's own ears, unencumbered. The condition is that large head movements are restricted [13]. So, with perfect HRTFs perceptions can be wrong.

2. It is possible to modify, in various ways, the acoustical performance of the external ears and, so long as head movements are possible, spatial perceptions remain remarkably stable. Years ago I encountered a product made by Acos. It consisted of an effective hard-shelled, liquid-filled-cushion ear defender, with headphone drivers inside the cups, and one-inch microphones on the outside of each one, facing left and right. In the headband were amplifiers that provided the person wearing the device with binaural communication to the outside world. It was unity gain until the sound level approached dangerous levels, and then the amplifier smoothly limited. The idea was brilliant; workers in noisy environments would have natural hearing except that very loud sounds would be limited to safe levels. The listener was functioning without external ears – no normal HRTFs – and the locations of the microphones expanded the inter-ear distance by over three inches. Since these were designed for hearing conservation and voice communication, fidelity was not great. Yet, after a very brief adaptation interval, even the most skeptical person functioned and navigated normally. There were no problematic front-back reversals or in-head localizations. But, normal head movements were present.

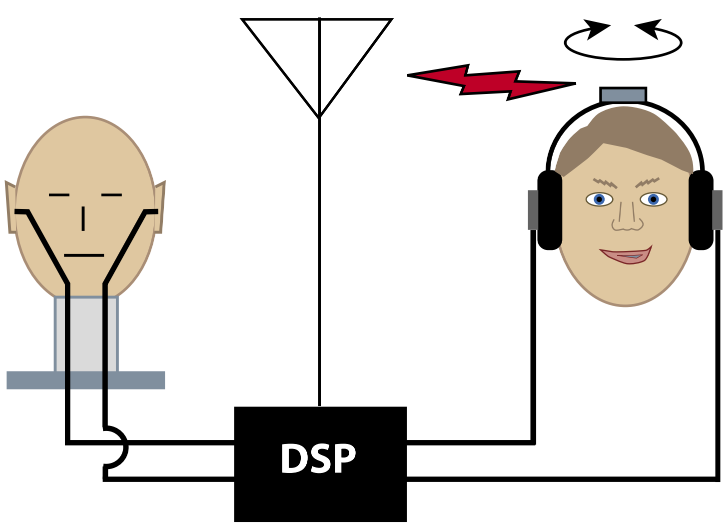

Figure 9. A binaural system in which head position is interpreted by a sensor (usually by magnetic fields) and appropriate real-time changes are made to the binaural signals. The listener gets the impression that the mannequin head follows his movements, even though the binaural signal may be a recording.

Looking at these two examples, one could argue that, if we are allowed head movements, errors in HRTFs are relatively unimportant.

Nowadays we have the means to do both: maintain basic HRTF information while allowing head movements, see Figure 9. The availability of devices that can track the position of a head in three dimen-

2. A head-related transfer function (HRTF) is measured as an impulse response (amplitude vs. time) or as the Fourier equivalent, amplitude and phase vs. frequency. Measurements are made at the eardrum location, from a sound source at some distance from the head. The nature of the HRTF is strongly individual, depending on the shape and size of the head and external ear (pinna), and it is a function of the vertical and horizontal angle of the sound source with respect to the ear. To characterize the directivity of an ear, a set of HRTFs is required, measured at many vertical and horizontal angles around the head.

sions, and digital processing that can modify binaural sounds in real time, has given us the means to crack this puzzle. Various head-tracked binaural systems have been built in recent years, including one that auralizes phantom loudspeakers in a listening environment as well [14]. We are likely to see much more of this in the future. Such systems clearly have application for computer games and in rear-seat entertainment installations in cars.

SYNTHESIZING BINAURAL SIGNALS

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. Of course this allows us to synthesize multichannel audio systems of almost any format.

To achieve this we must know, for every direction we wish to synthesize, and 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 (HRTFs) to each of the ears. With this information stored away, the binaural directional synthesizer can convert any single-channel signal into the binaural left- and right-ear signals appropriate to the chosen direction, see Figure 11.

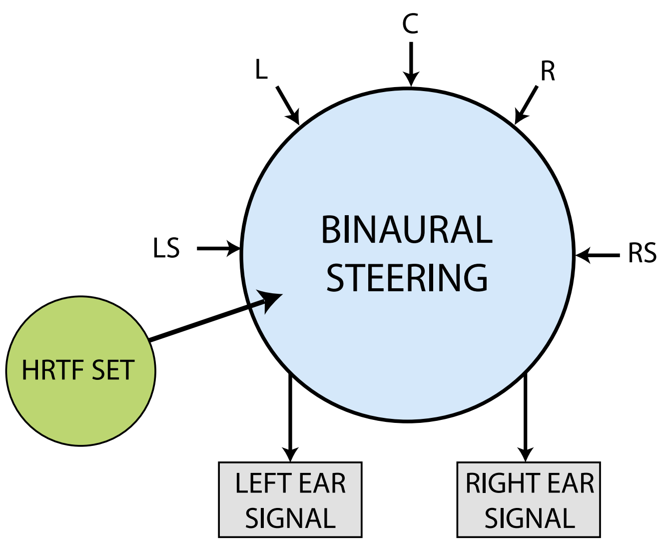

Figure 10. By convolving input signals with the appropriate left- and right-ear HRTFs it is possible to create the binaural signals necessary to cause the images to be perceived at predetermined locations in space. Reproduction can be through headphones or crosstalk-cancelled loudspeakers. Here we are creating a 5-channel phantom home theater with which to enjoy surround sound.

This process has been widely used since about 1980, when affordable computers of sufficient speed and power were available. A lot of the early effort was stimulated by military and aerospace applications, particularly the notion of delivering 3-D auditory imaging to fighter pilots to help them cope with information overload in a battle situation. If voice communication with friendlies and radar information about enemies could be directionally encoded, it would help greatly with the important issue of maintaining a three-dimensional perspective on what is happening. Durand Begault provides background on this subject in his book [18].

For do-it-yourselfers, the MIT Media Lab has made available a set of HRTFs for a KEMAR mannequin on their web site (http://sound.media.mit.edu/KEMAR.html).

Putting all of the ingredients together, it should be possible to create credible illusions of externalized, unambiguously front/back localized, phantom loudspeakers, through which to reproduce multichannel audio. This requires the head-tracking technology shown in Figure 9, combined with the binaural steering technology of Figure 10. To complete the auditory picture, it is possible to incorporate the acoustical ambience of a particular room, so that the multichannel playback appears to take place in a familiar, real, acoustical environment [14], Figure 11.

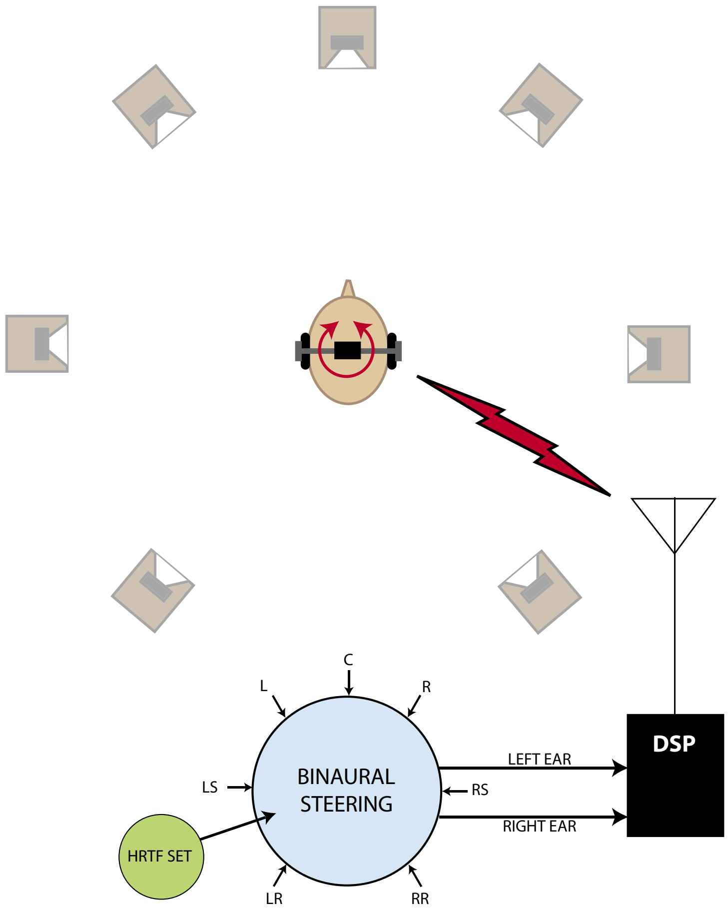

Figure 11. The ultimate headphone home theater simulation? Head position tracking is combined with binaural digital processing so that the perceived directions of sounds remain fixed in space as the head rotates. Frontal sounds are correctly externalized, at a distance. In-head localizations and front/back reversals are eliminated. A binaural steering algorithm creates the left- and right-ear signals corresponding to the selected angular locations of some number of surround channels. Here seven are shown, but more are possible including, elevation channels. The headphones depicted here have a wireless connection to the base station, thus avoiding the restriction of a tether, and the annoyance of mechanical wire noises. In conventional headphone listening the problem is that the orchestra or band moves with the head. Here this is gone, and we have the intriguing situation of the listener being able to take a multichannel audio system and the listening room to any location. This feature goes beyond simple entertainment, because it makes such a scheme attractive to professionals for the monitoring of remote recordings.

CROSSTALK-CANCELLED BINAURAL AUDIO

There is another way to ensure that the sound sources are convincingly external and in front – reproduce the binaural signals through loudspeakers. The problem now 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. Without independent signals at the ears, we do not have binaural audio, we have stereo.

But, 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, see Figure 12.

The crosstalk-cancellation concept was patented by Atal and Schroeder at Bell Telephone Laboratories in 1966 [15], 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.

Simplified 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], Polk's SDA loudspeakers [16], and the "panorama" mode in Lexicon's surround processors.

In the 1980s 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 [17]. 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 fifty-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, works very well. At greater distances one must pay attention to minimize reflected sounds, which can be done by controlling the directivity of the loudspeakers or by applying absorption to 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.

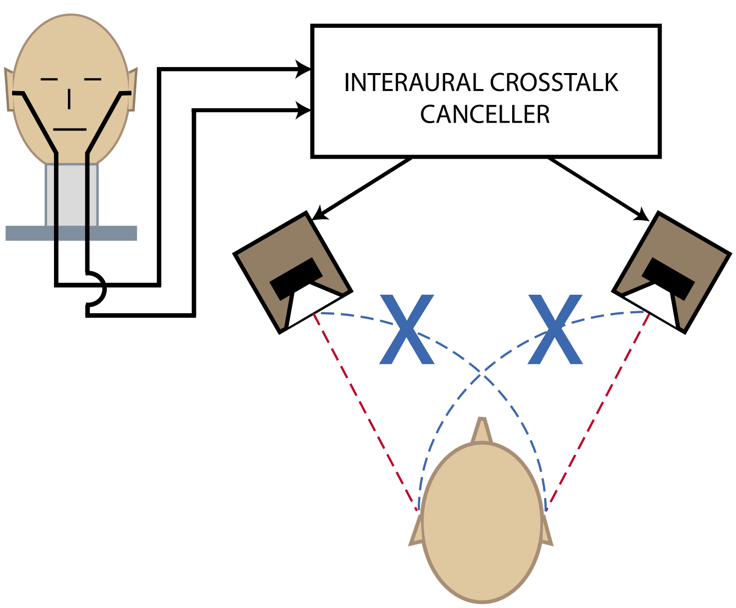

Figure 12. A binaural system in which the signals are electronically processed in a manner that results in a cancellation of the acoustical crosstalk when all of the sounds arrive at the ears. The left loudspeaker then communicates only to the left ear and the right loudspeaker communicates only to the right ear. The crosstalk cancellation is not perfect but, it can be extremely effective. Obviously, the cancellation processing only works when the listener is exactly in the sweet spot, and different processing is necessary for different loudspeaker angular arrangements relative to the listener.

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


AKG HEARO 999 Audiosphere II

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. [11][12][13]

THE NATURE OF THE "SWEET SPOT"

In *any* 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 scalable. 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.

An important challenge in all of this is to maintain sound quality – fidelity– while performing these manipulations. Sadly some systems not only distort the spatial imagery, they distort the sound as well.

"fuzzy" sweet spot = "fuzzy" localization

SPEAKER-BASED BINAURAL AUDITORY EFFECTS

1. SPEAKER SPREADER.

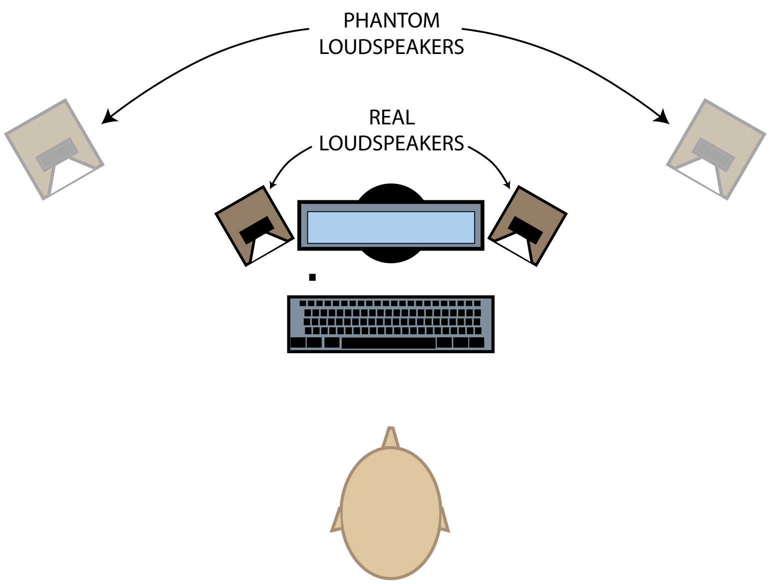

Figure 13. With crosstalk cancellation and binaural steering combined, it is possible to position phantom loudspeakers which can then be used to reproduce conventional stereo. The angular separation could be decided by the listener.

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.

A system of this kind produced some of the most persuasive stereo imaging the author has ever heard.

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. The soundstage is enormously wide, ranging ±90°, if you happen to like that kind of thing. 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, as shown in Figure 14. 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, in the absence of a head-position tracking system, the frontal soundstage may be less than convincingly external. However, with the visual image of a movie, this tends to disappear. AKG has various versions of this available in wireless headphone implementations.

If we take the next step and create a crosstalk-canceled loudspeaker version, then the listener can have a multichannel experience 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.

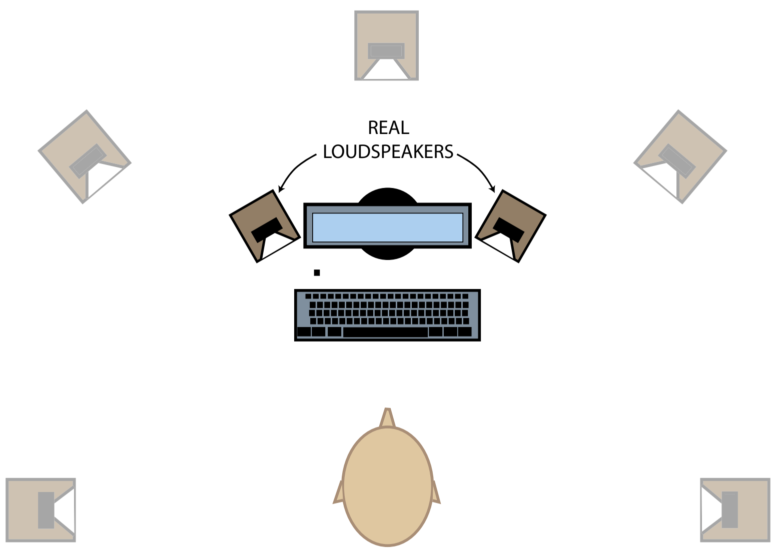

Figure 14. Binaurally-steered phantom loudspeakers can be used to reproduce the five channels from any multichannel format.

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.

Some systems in the marketplace, masquerading as phantom home theater, cannot create the five independent phantom sources. Playing test signals through the system, in which each channel is driven separately in sequence, instantly reveals such systems. Seated in the sweet seat, one should hear five quite separate phantom loudspeakers being activated.

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 determined by the quality of the sounds arriving at our ears. Conventional stereo and multichannel audio 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 variable in the critical final phase of the delivery system. In fact, there are so many uncontrollable variables that all sound reproduction systems are different in some way or other.

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 that also fits nicely with audio/video workstations, which are becoming increasingly common in production and post-production tasks in film, television and, of course, multimedia and computer games. For these situations, crosstalk-cancelled binaural audio is a natural.

Headphone reproduction is commonly associated with in-head localization (with stereo signals) and with a lack of convincingly externalized sound images in the frontal region (with binaural signals). However, a head-tracking adapter changes all of that. Add some listening room ambience, pay some attention to ensuring that the headphones are delivering the correct spectrum to the eardrums, and the overall experience is radically transformed. Suddenly, the experience is not one with compromises, but one with unique advantages.

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 binaural systems, all of the sound comes from the same two transducers 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. In binaural reproduction, the listener has none of the traditional information by which to judge the distance of the real sources of sound. Impressions of distance then are derived from cues that are in the recordings. It is intriguing to be aware 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 headphones, or 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.

At two large trade shows VMAx Phantom Home Theater was demonstrated. It was compared with a conventional five loudspeaker implementation, using identical loudspeakers everywhere. Many people experienced the comparison and, when asked afterwards which system they preferred, about half had no preference. That alone was gratifying, but even more encouraging were the 20% who preferred the phantom theater. Why? The seamlessness of the soundstage and the impressive sense of depth. One could always aware of the five real speakers, whereas the phantoms drew less attention to themselves.

Will it replace five speaker surround? No. It is an antisocial experience for one thing, and if one has a big screen in a big room, and several listeners, multichannel is the way to go. But there are situations when …

TRUE SUCCESS IS IN THE ATTENTION TO DETAILS

For loudspeaker-based binaural, convincingly good spatial effects require 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 naïve. Disappointments will abound. Fortunately, with attention to the right details, it is not difficult to be very successful.

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

For headphone-based binaural, there must be some assurance that the sounds arriving at the eardrums are as they should be. The belief that headphones are inherently higher fidelity than loudspeakers is incorrect. They differ widely, and some common designs exhibit different performance with minor changes in position relative to the ears, so just taking them of and putting them on introduces variability. Some kind of calibration may be required.

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 HRTFs 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

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 times and places for both kinds. Technology gives us a choice.

We knew long ago that there were better ways. As a stopgap (a fifty-year one!) stereo has been very enjoyable. It is time to move on, though. Are you on board?

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. Griesinger, D. , "Multichannel Matrix Surround Decoders for Two-Eared Listeners", 101st Convention, Audio Eng. Soc. Preprint No. 4402, Nov. 1996.

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

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

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

14. Horbach, U., Karamustafaoglu, A., Pellegrini, R., Mackensen, P., Theile, G. "Design and Applications of a Data-Based Auralization System for Surround Sound" 106th Convention, Audio Eng. Soc., Preprint 4976, April 1999

15. 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).

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

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

18. Begault, Durand, "3-D Sound for Virtual Reality and Multimedia", AP Professional, 1994.

19. See HEARO products at http://www.akg.com.

This paper is based on a two-part article entitled "The Future of Stereo" which appeared in the May and June 1997 issues of Audio magazine.

JBL Infinity harman/kardon REVEL lexicon AKG

www.jbl.com www.infinitysystems.com www.harmankardon.com www.revelspeakers.com www.lexicon.com www.akg.com