

---

In any case, other systems exist that do a better job where large distances or low frequencies are concerned. For example, the study of thunder signatures requires an array (four or more) of spaced microphones to build a three dimensional representation of them. Wildlife habitats, also, can be studied by monitoring the playback of four or more widely spaced microphones. The main problem here is that sounds very near to the microphones tend to mass, in playback, around the corresponding loudspeaker; conversely, low level sounds emanating from the area equidistant from the microphones get largely lost.

## **Binaural Systems**

Two systems are here proposed which endeavor to retain the benefits of binaural arrangements, but, in a four-channel configuration. And, like Ambisonics, they will do so in such a way that the entire soundfield can be later matrixed for either 360° two-dimension, or three-dimensional spatial resolution.

The first system consists of four omnidirectional capsules, configured in a plane parallel to the ground, so that the distance between any pair of capsules is about 16 cm., or the spacing between a pair of typical human ears. Further, these capsules are separated by a baffle that casts an acoustic shadow, so that the distance between capsules that the sound must travel is actually more like 32 cm. This is a typical distance between the ears, taking into account the head shadow.

This system is adequate for resolving the majority of sounds interesting to humans (large mammals have little to fear from the sky, so height resolution is not so critical as it is for, say, a mouse, which must be constantly aware of what—e.g., a hawk—is above it). But, where height resolution is necessary, a variant is proposed, where the capsules are configured in a tetrahedral array, but with the same inter-capsule dimensions.

While it is true that neither system can emulate the pinnae effects of a true binaural system, it is felt that this lack can be largely compensated for in software. In this regard, it is envisioned that both systems would employ, for post-processing, DSP hardware, for which algorithms can be more easily implemented and altered than is the case with Ambisonics.

---

## Four-Channel Systems

On the other hand, a full soundfield can be represented accurately to a more realistic rotating-head model with as few as four channels. Further, the simulated reality is interactive to the extent that the listener can pick a view, any view, as desired. The concept of employing four channels to present an actual or a virtual soundfield has been around at least as far back as Alan Blumlein's research in "Ambisonics" for the BBC in the 1930s, and had a brief fling with consumer popularity in the 1970s, when quadraphonics became somewhat of a buzzword. Unfortunately, the implementation died in infancy, being ill-conceived, both technically and commercially, if not linguistically—as Glenn White points out, "any system that mixes Latin and Greek to invent its name is probably doomed from the start."

However, the idea has, in the last few years, been significantly reincarnated as "surround sound", specifically as an adjunct to film and video presentations. As this industry is a substantial one in terms of commercial potential, it may be well to consider the various four-channel schemes for original recording.

### Ambisonics

The Soundfield Microphone™, built by Calrec (now AMS), uses one "single-point" device, which actually consists of one omnidirectional and three figure-of-eight microphones all sharing (as closely as possible) a single point in space. Through a complex bit of matrixing (described by Michael Gerzon), the signals can be made to represent the soundfield that exists at the position of the microphone at the time of recording. The system has been improved since its original implementation. Nevertheless, it has never enjoyed a widespread success, owing in part to its expense, but also to its reliance on cumbersome and fussy peripheral hardware, and especially to the fact that a full implementation requires a recording format with four channels (heretofore not available for portable applications), and matrixing four into two is not completely successful. Another drawback is that such systems do not translate well to binaural playback, as there are no IATDs among the capsules of a single-point microphone device. Still, the steer-ability of this system suggest that it is quite useful when modelling a rotating-head.

### Spaced Arrays

It can be argued, of course, that a rotating head model suggests a reality only slightly more natural than those with fixed heads, as the listener is still unable to move through space. But, in fact, *many* selectable views from one position represent an element of control over apparent reality that is entirely non-existent with *one* view from that same position. (To be sure, there is a world of difference between *real* reality and a simulation of it; it is possible to render a cup, say, in two or even three dimensions with a life-like hologram of it, but one is not allowed to drink from it.)

---

## Appendix

### Concept Proposal for Four-Channel Soundfield Systems

As probability and chaos theoreticians are wont to point out, reality is a complex bit of business—so much so that its elements are, quite literally, impossible to determine with precision. Weather, for instance, can not be accurately predicted without an pretty much infinite number of sensors. Neither can an acoustic *reality* (i.e., the scoundscape, that is both interactive and intermodal) be wholly represented by a finite number of transducers.

Therefore, audio, insofar as it attempts to represent an acoustic reality, is, and must remain, an art of illusion. Not that this is bad: The brain is able, through some rather complex processing schemes, able to extrapolate, from minimal cues, soundfields that have never existed, so is able to model potential and archetypal experiences—clearly an aid to survival, at least. As a direct analog of this, acoustic data can, through the use of clever software, be coerced into presenting a very credible soundfield.

Data to represent such soundfields can be recorded, more or less satisfactorily, by a variety of two-channel techniques, each of which has its pluses and minuses:

- Spaced pairs, for instance, provide a sometimes startling sense of space, enhanced by the fact that the signals between them remain uncorrelated at quite low frequencies (leading to a larger-than-life feeling). However, directional localization is problematic, as the only available cues are IATDs and spectral differences between the capsules.
- Coincident techniques provide excellent resolution left-to right, but, as they are entirely lacking in IATDs, they have trouble with the height and depth dimensions. Also, since coincident pairs utilize directional capsules, the low frequencies tend to get lost, further diminishing spatial resolution.
- Near-coincident pairs also suffer from the latter malady, although they often represent the best compromise when some sense of space is desired while retaining good lateral resolution.
- Binaural and quasi-binaural systems can sound uncannily real in all dimensions, at least with headphone playback. This is because they represent to a high degree the true presentation of the soundfield upon the typical ear positions. But they, along with all two-channel systems, assume a fixed head listening model; since the microphone positions are fixed, relative to both the soundfield and to each other, the listener has no possibility of altering his or her own “view” of the space presented. As noted in the body of the paper, this is a rather unnatural state of affairs.

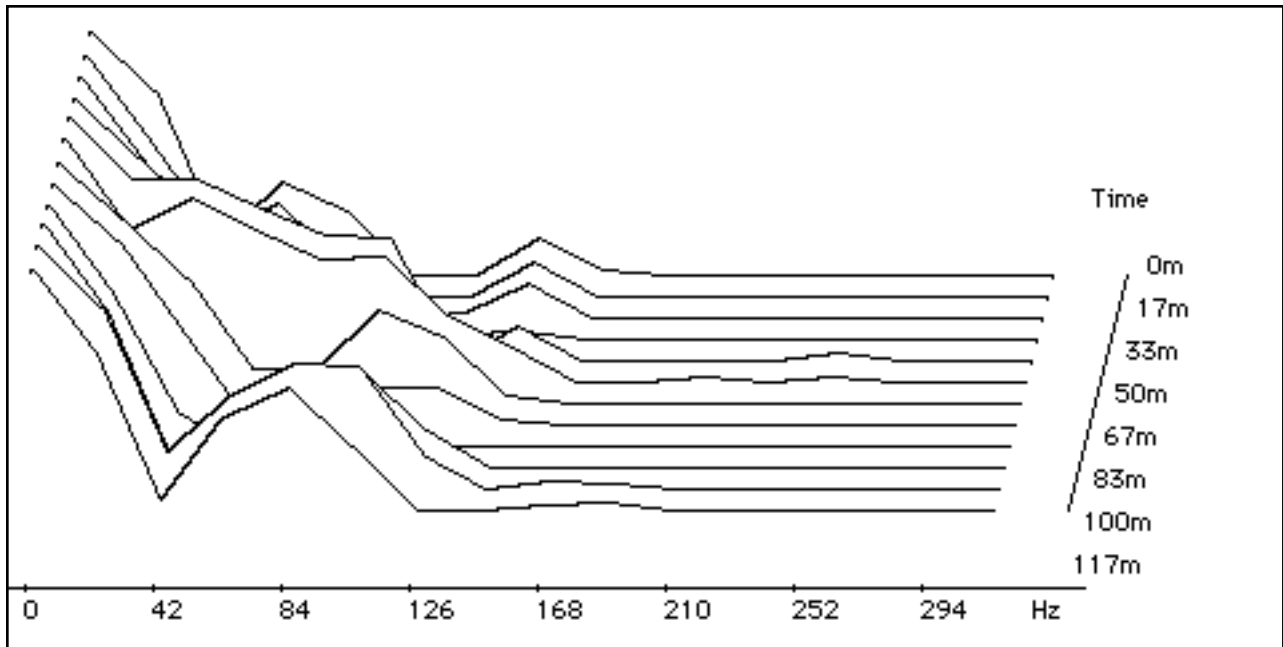


Fig. 4f. At 30 seconds.

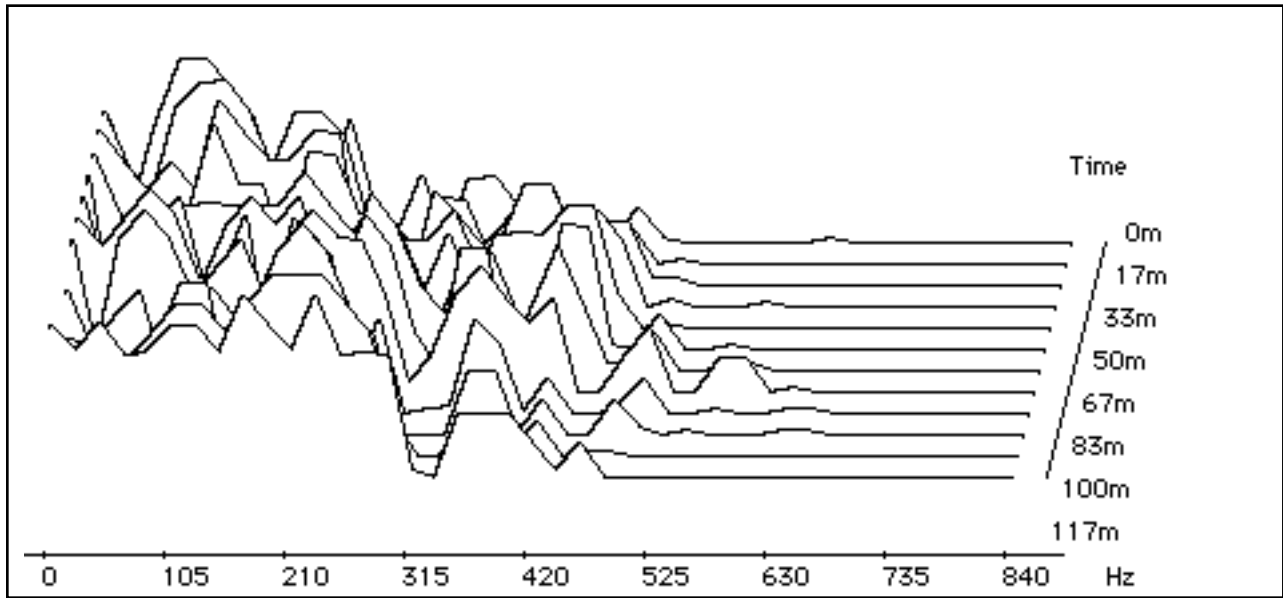


Fig. 4d. At 10 seconds.

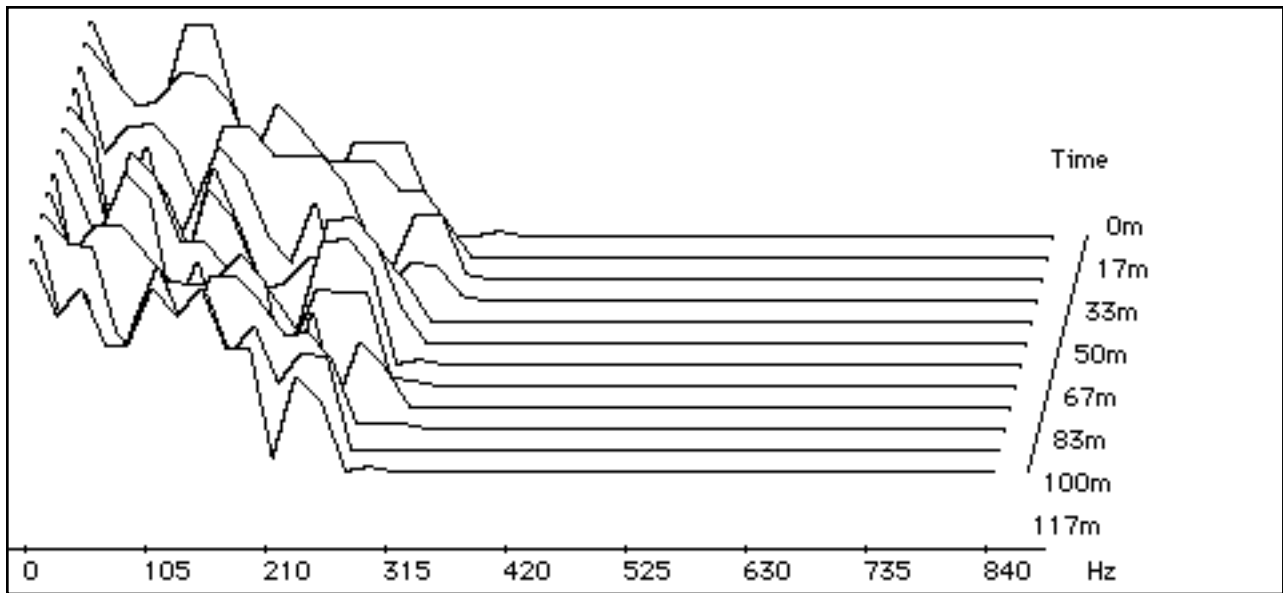


Fig. 4e. At 20 seconds.

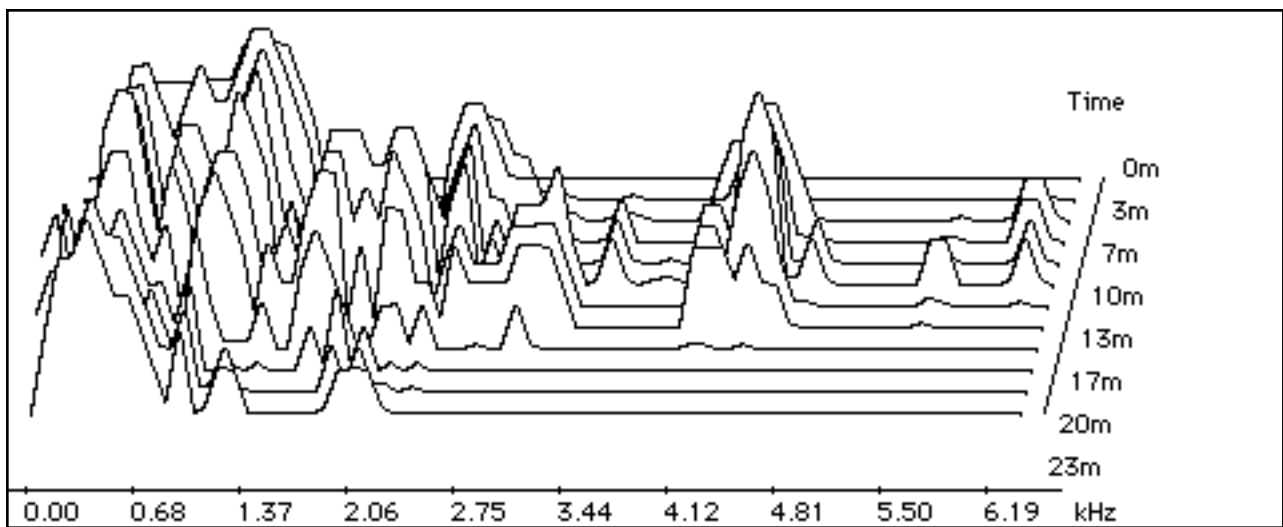


Fig. 4a. FFT of the first 20 msec. of the impulse.

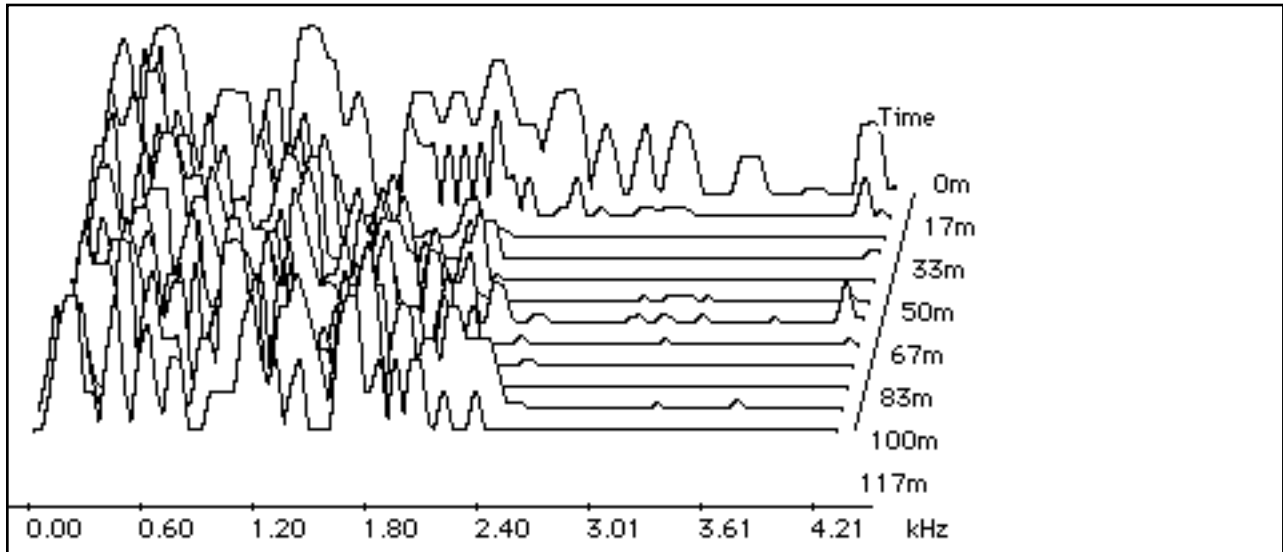


Fig. 4b. The first 100 msec.

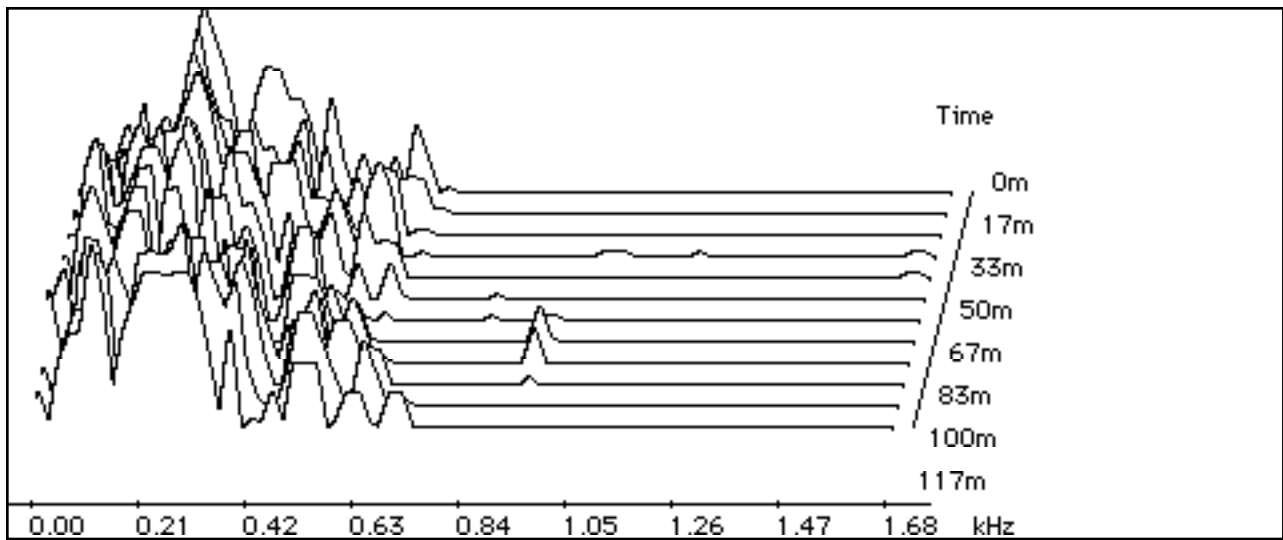


Fig. 4c. At 5 seconds

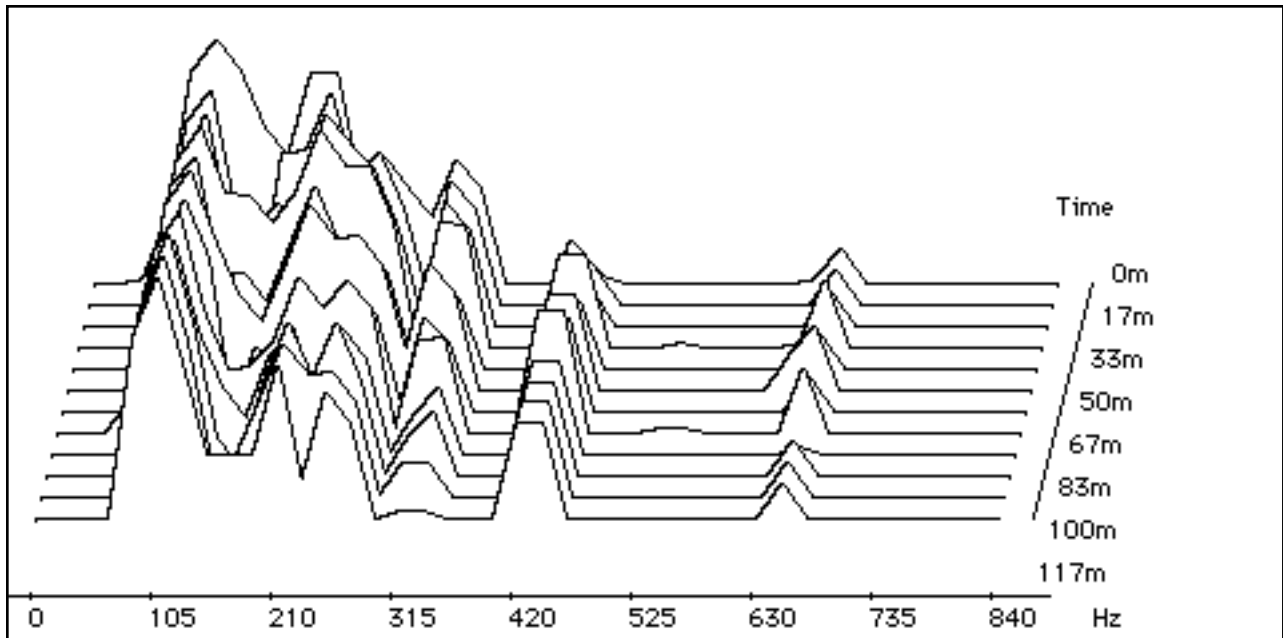


Fig. 2. Guitar *tambora* effect FFT. Note the obvious harmonic structure, especially the lack of upper harmonics. But the impulse excites the acoustic space in complex, and not always predictable, ways.

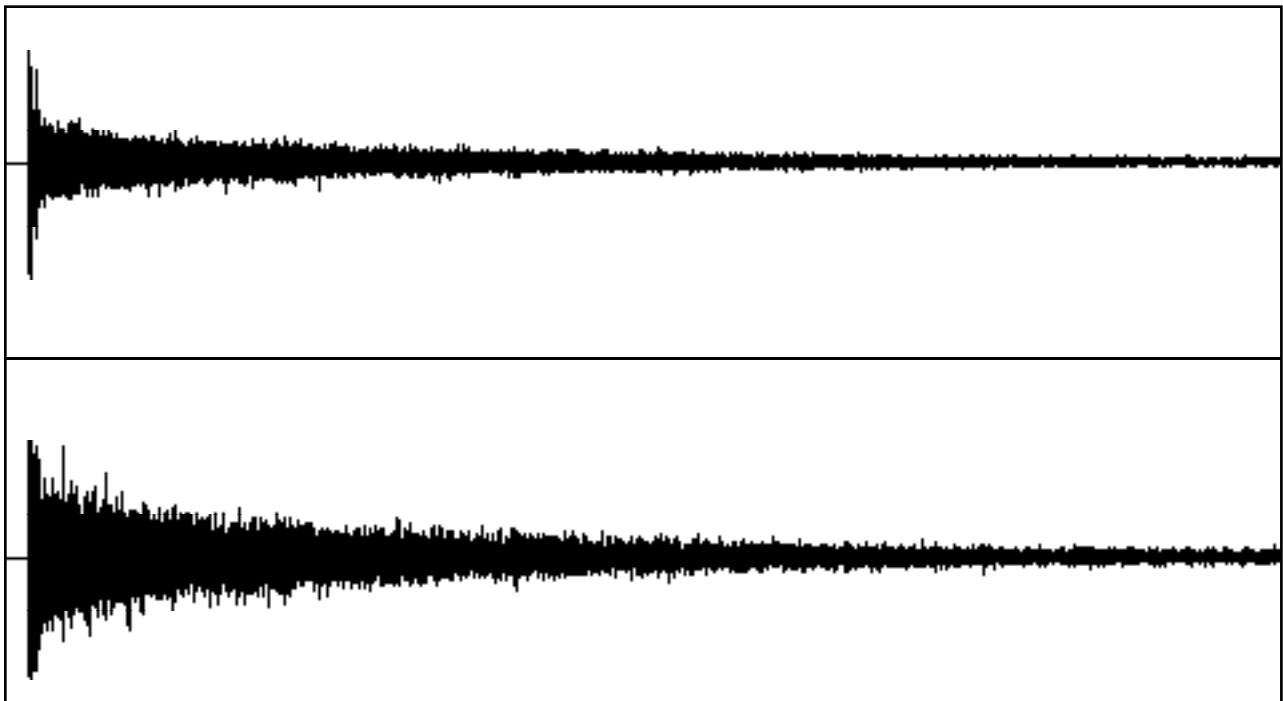


Fig. 3. Two channel envelope plot of an acoustic impulse (in this case, the “pop” of a balloon) in the cistern. Only the first 12 seconds are plotted here, but the envelope retains similar characteristics for a minute or more.

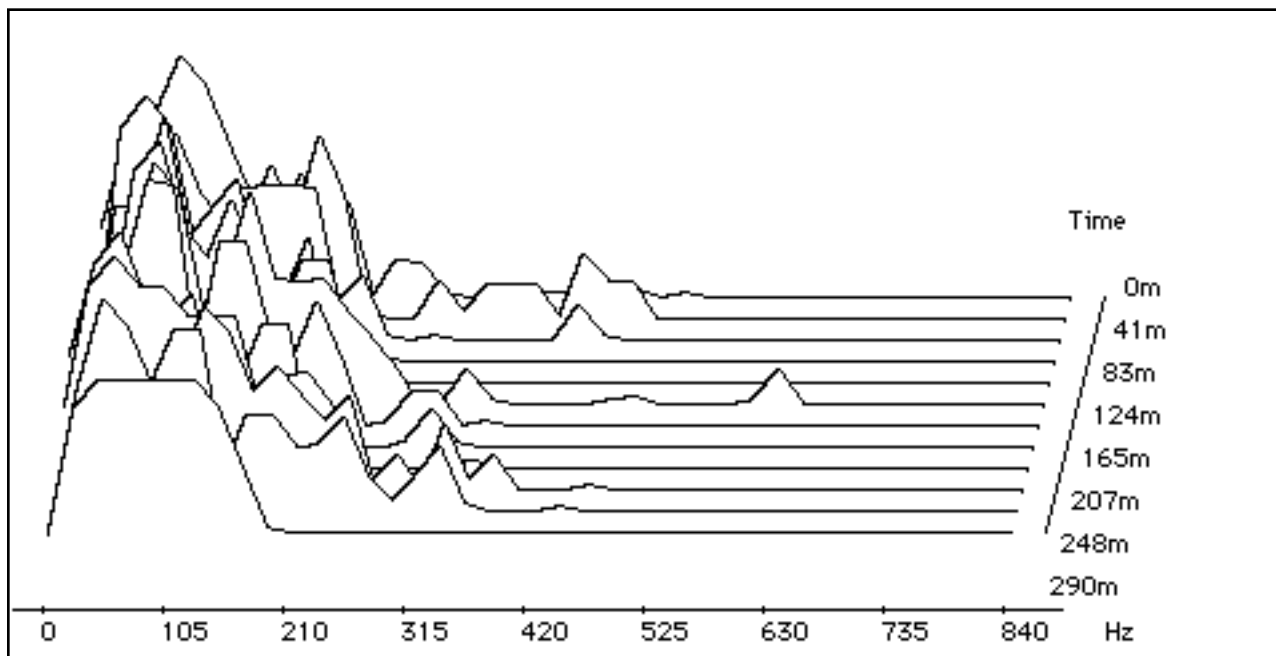


Figure 1b. Thunder roll at 20 seconds. By this time, most high frequencies have become heavily attenuated.

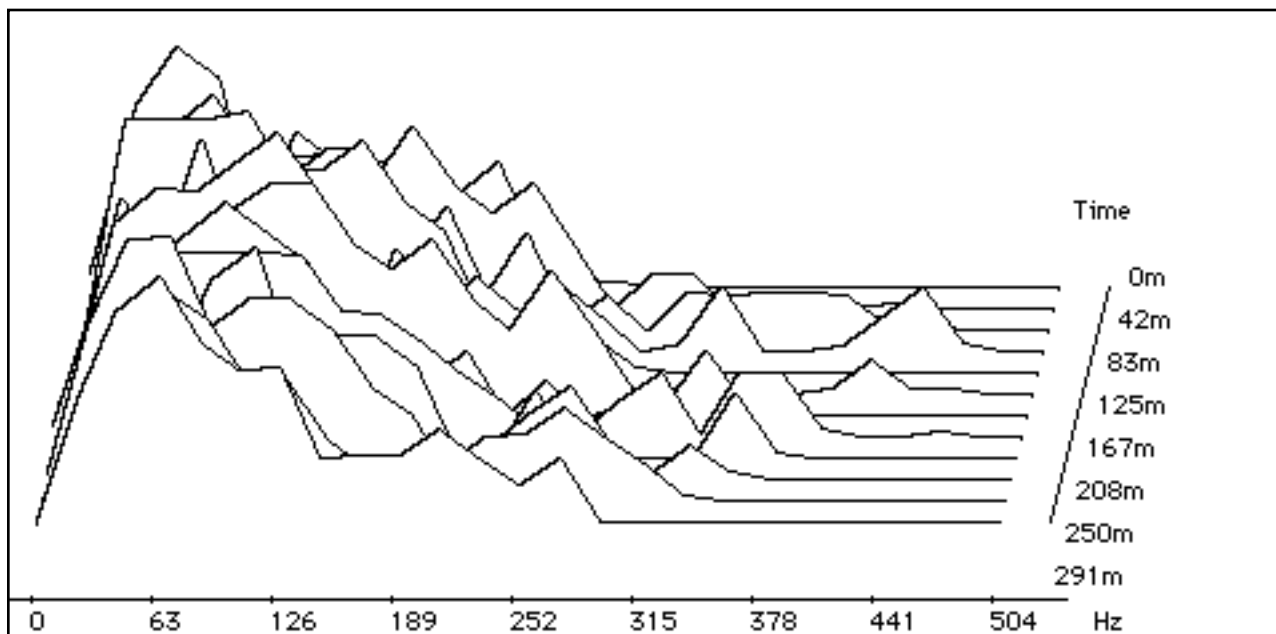


Fig. 1c. Thunder roll at 30 seconds. At this point, there is almost no high frequency information radiated by the thunder available at the microphones. The few components seen are from extraneous sounds.

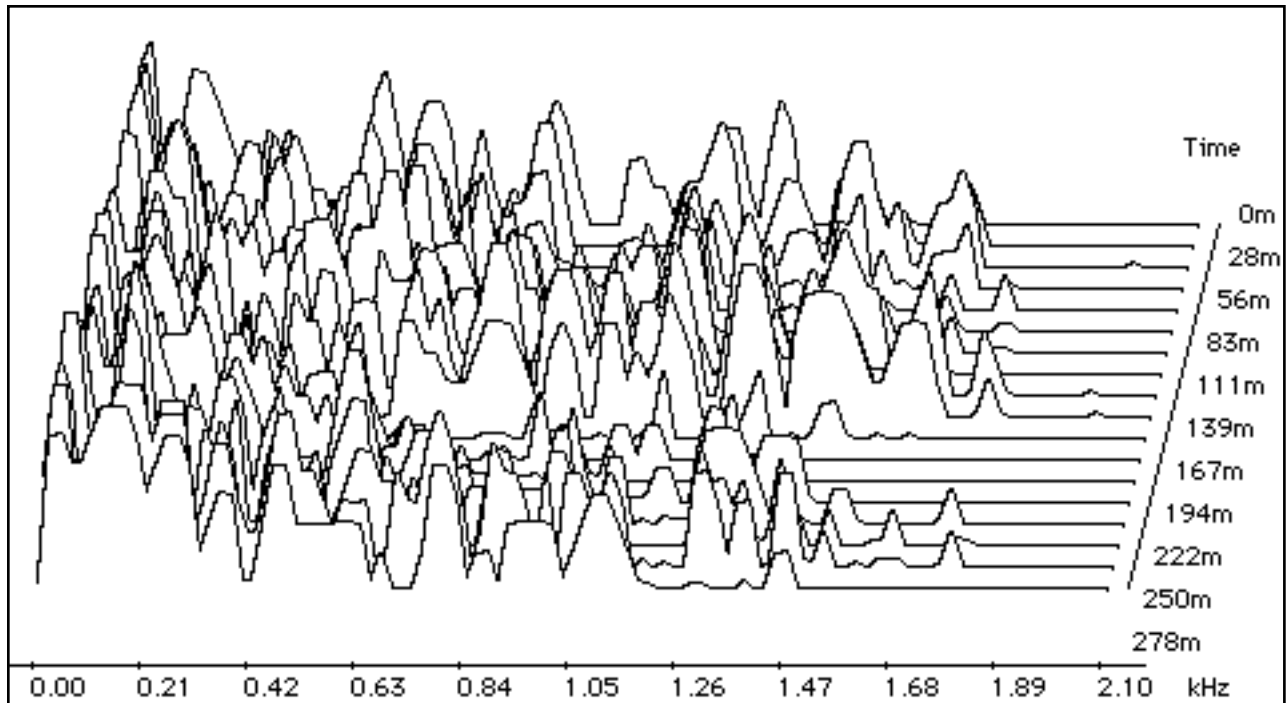


Figure 1a. FFT of thunder pre-clap pulses. (All FFT displays were normalized, and so do not show relative amplitudes sequentially.)

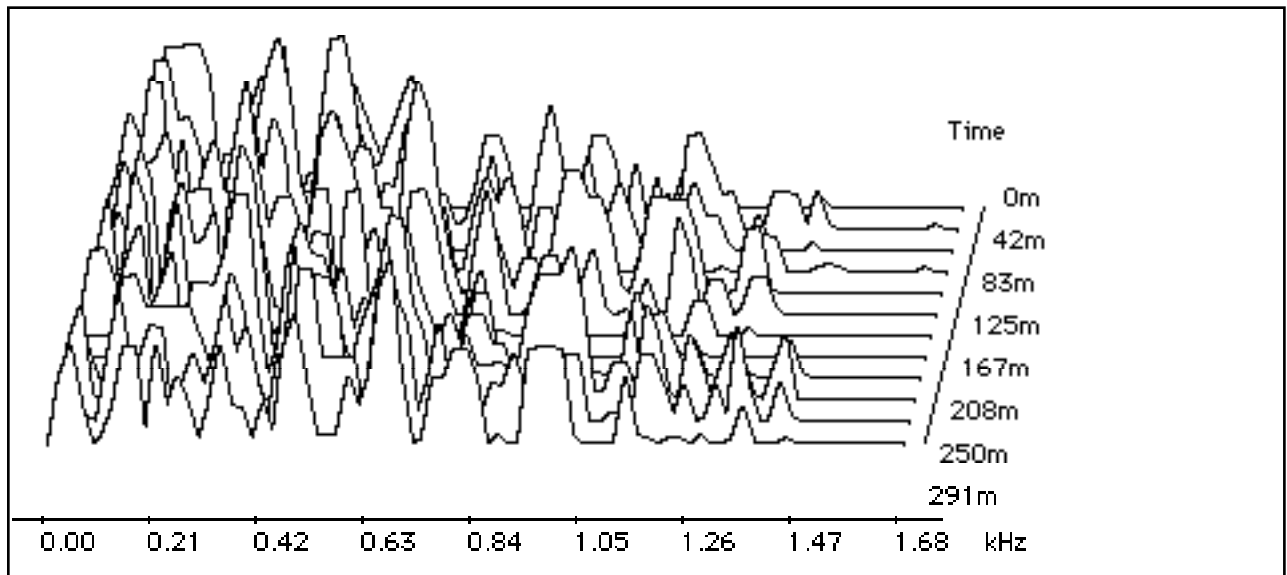


Fig. 1b. Thunder clap. The lightning channel for this thunder was quite close, as can be seen by the relatively high amount of high frequency information. Subjectively, a thunder clap of this proximity results in a “ripping” sound.

- 
- [31] D. Ackerman, *A Natural History of the Senses* (Random House, New York, 1990).
- [32] W.C. Stebbins, *The Acoustic Sense of Animals* (Harvard University Press, Cambridge, MA, 1983).
- [33] G. Martin, *Birds by Night* (T & AD Poyser, London, 1990).
- [34] B.H. Lopez, *Of Wolves and Men* (Charles Scribner's Sons, New York, 1978).
- [35] D.M. Guss, *To Weave and Sing* (University of California Press, Berkeley, 1989).
- [36] C. Turnbull, *The Mountain People* (Simon and Schuster, New York, 1972).
- [37] L. van der Post, *The Heart of the Hunter* (Harcourt Brace Jovanovich, San Diego, 1961).
- [38] C.F. Jopling (Ed.), *Art and Aesthetics in Primitive Societies* (E.P. Dutton, New York, 1971).
- [39] G. Thiele, "Further Developments of Loudspeaker Stereophony," AES Preprint # 2947, presented at the 89th AES Convention, Los Angeles, September, 1990.
- [40] R. Streicher and D. Dooley, "Basic Stereo Microphone Perspectives—A Review," *J. Audio Eng. Soc.*, vol. 33, pp548-556 (July/Aug., 1985).
- [41] G.D. White, *The Audio Dictionary* (University of Washington Press, Seattle, 1987).
- [42] J. Eargle, *The Microphone Handbook* (Elar Publications, Plainview, NY, 1981).
- [43] A.A. Few, "Thunder Signatures," *Trans. Am. Geophysical Union*, vol. 75, pp. 7517-7523 (May, 1974).

- 
- [16] K. Genuit, H.W. Gierlich, and W. Bray, "Development and Use of Binaural Recording Technique," AES Preprint #2950, presented at the 89th AES Convention, Los Angeles, September, 1990.
- [17] M. Wöhr, G. Theile, H. Goeres, and A. Persterer, "Room-Related Balancing Technique: A Method for Optimizing Recording Quality," J. Audio Eng. Soc., vol. 39, pp. 623-631 (Sept., 1991).
- [18] H.W. Gierlich and K. Genuit, "Processing Artificial-Head Recordings," J. Audio Eng. Soc., vol. 37, pp. 34-39 (1989, Jan./Feb. 1989).
- [19] D.R. Begault, "Challenges to the Successful Implementation to 3-D Sound," J. Audio Eng. Soc., vol. 39, pp. 864-870 (Nov., 1991).
- [20] E.A. Cohen, "Technologies for Three Dimensional Sound Presentation and Issues in Subjective Evaluation of the Spatial Image," AES Preprint # 2943, presented at the 89th AES Convention, Los Angeles, September, 1990.
- [21] E. Macpherson, "A Computer Model of Binaural Localization for Stereo Imaging Measurement," J. Audio Eng. Soc., vol. 39, pp. 604-622 (Sept., 1991).
- [22] H. Shapiro, 1962 Forward to Colin Turnbull, *The Forest People* (Simon and Schuster, New York, 1961).
- [23] S.P. Lipshitz, "Stereo Microphone Techniques: Are the Purists Wrong?" J. Audio Eng. Soc. (Features), vol. 34, pp.716-44 (Sept., 1986).
- [24] D.H. Cooper, "Problems with Shadowless Stereo Theory: Asymptotic Spectral Status," J. Audio Eng. Soc., vol. 35, pp. 629-642 (Sept., 1987).
- [25] A.D. Blumlein, British Patent 394,325 (June, 1923); reprinted in J. Audio Eng. Soc., vol. 6, pp. 91-98 (April, 1958).
- [26] G. Thiele, "On the Standardization of the Frequency Response of High-Quality Studio Headphones," J.. Audio Eng. Soc., vol. 34, pp. 956-969 (Dec., 1986)
- [27] B. Burt, cited in H. Tomlinson, "New Factors in Sound for Cinema and Television," J. Audio Eng. Soc., vol. 39, pp. 529-539 (July/Aug., 1991).
- [28] O.B. Hardison, Jr., *Disappearing Through the Skylight* (Penguin , New York, 1989).
- [29] J. Lackner, "Influence of Posture on the Spatial Localization of Sound," J. Audio Eng. Soc., vol. 31, pp. 650-661 (Sept., 1983).
- [30] B. Mandelbrot, *The Fractal Geometry of Nature* (W.H. Freeman, San Francisco, 1983).

---

## References

- [1] J.C. Steinberg and W.B. Snow, "Auditory Perspectives—Physical Factors" *Electrical Engineering*, vol. 53, pp. 12-15 (Jan., 1934).
- [2] J. Eargle, *The Recording Engineer's Handbook* (Von Nostrand Reinhold, New York, 1992).
- [3] Y. Ando, *Concert Hall Acoustics* (New York, Springer Verlag, 1985).
- [4] G. Von Békésy, *Sensory Inhibition* (Princeton, Princeton University Press, 1967)
- [5] W.B. Snow, "Basic Principles of Stereophonic Sound," *J. Soc. Motion Picture and Television Engineers*, vol. 61, pp. 567-589 (Nov., 1953).
- [6] J.P. Blauert, *Spatial Hearing* (M.I.T. Press, Cambridge, MA , 1983).
- [7] B. Bartlett and M. Billingsly, "An Improved Microphone Array Using Boundary Technology: Theoretical Aspects," *J. Audio Eng. Soc.*, vol. 38, pp. 543-552 (July-Aug., 1990).
- [8] G. Thiele, "On the Naturalness of Two-Channel Stereo Sound," *J. Audio Eng. Soc.*, vol. 39, pp. 761-767 (Oct., 1991).
- [9] M. Gerzon, "Ambisonics in Multichannel Broadcasting and Video," *J. Audio Eng. Soc.*, vol. 33, pp.859-871 (Nov. 1985).
- [10] D.H. Cooper and J.L. Bauck, "Prospects for Transaural Recording," *J. Audio Eng. Soc.*, vol. 37, pp. 3-19 (Jan./Feb. 1989).
- [11] D. Griesinger, "Equalization and Spatial Equalization of Dummy-Head Recordings for Loudspeaker Reproduction," *J. Audio Eng. Soc.*, vol. 37, pp. 20-29 (1989, Jan./Feb. 1989).
- [12] W.R. Woszczyk, "A New Method for Spatial Enhancement in Stereo and Surround Recording," AES Preprint No. 2946, presented at the 89th AES Convention, Los Angeles, September, 1990.
- [13] H. Møller, "Reproduction of Artificial-Head Recordings through Loudspeakers," *J. Audio Eng. Soc.*, vol. 37, pp. 30-33 (1989, Jan./Feb. 1989).
- [14] D.J. Furlong and A.G. Garvey, "Spaciousness Enhancement of Stereo Reproduction using Spectral Stereo Techniques," AES Preprint #3007, presented at the 89th AES Convention, Los Angeles, September, 1990.
- [15] K. Kotorynski, "Digital Binaural/Stereo Conversion and Crosstalk Cancelling," AES Preprint #2949, presented at the 89th AES Convention, Los Angeles, September, 1990.

---

## Conclusions

As we explore the above examples, most of which are not “mainstream” (music), a philosophical question becomes pertinent: Why record *other* than “mainstream” (audio)? After all, there is, they say, glamour and lots of money in studio work, and prima donna artistry aplenty in large-scale classical recording. Sound effects are definitely a niche market, and the wackier stuff is best done on synthesizers, right?

Perhaps these things are so. And yet, is our species, or at least our ostensibly civilized part of it, so far removed from the day-to-day concerns of survival that we can afford to ignore the development of our *seuketat*, an Eskimo term meaning “ear of the animal”? Sentience is transient and hauntingly ephemeral for all of us.

Writing, acting, composing, recording the music of the World: Following any worthy Muse demands a constant evaluation and re-evaluation of both her art and her craft. It is hoped that any instigations afforded by the present inquiries represent only the beginning of discovery.

## Acknowledgments

The author would like to thank Gordon Hempton, The Soundtracker™, and Jonathan Storm, EARTH TUNES™, for their invaluable comments and suggestions, and also the contribution of some of the examples.

---

a listener, and, except for very near positions to the source (wherever that might be), are simply overwhelmed by a cascade of diffusion. Since the reflections (all “near”, at least on the first pass) come from all locations in almost equal intensity, exact localization is extremely difficult to determine (indeed, for the typical listener, any kind of imaging, directional, spatial, or temporal, soon ceases to be of any concern). Also, the spectral character is heavily biased in favor of the low end (See Figure 4.): Over a 45 second or longer length of time, sound waves travel an appreciable distance, and air resistance tends to attenuate the high frequencies more than the low. Further, group delays can be considerable, with the higher frequencies in the indirect paths arriving up to several thousand degrees late relative to their associated low frequencies. This causes an extreme smearing of transients. Finally, given freedom of movement, the musicians can participate from several different locations during a piece.

In the absence of any prescribed aesthetic, two configurations were employed. The first used four omnidirectional microphones more or less equally spaced around the perimeter of the cistern in a quadraphonic configuration; the second used two of these microphones mixed with two near-coincident cardioid microphones about three meters from the “home base” of the instrumental performers. This latter scheme was eventually chosen for the commercial recording, as it allowed some clarity of direct sound, without which key clicks and other instrument noises were obliterated by diffusion. Interestingly, no dual-perspective qualities were discerned.

In the subsequent session, a third, binaural, configuration was included, and became part of a later commercial release. The principal advantage of the artificial head, was, as it turned out, not an acoustic one at all, since the human hearing has no clear-cut aesthetic preference in this environment, but a practical one: Since it was easy to move to a variety of locations in the cistern, many experiments could be undertaken. Also at this session, a technique was tried where microphones were placed within inches of the cistern’s outer wall. This resulted in capturing a certain amount of “spininess”, apparently an acoustic equivalent to the “wave” cheer performed in circular arenas, which had been previously noted.

The determination of an appropriate perspective was problematic, partly for the above reasons, but also for reasons of “staging”: The musicians spread themselves out widely (relative to the other musicians, that is), and turned and moved regularly. Again, however, the desire to find a perspective quickly became unimportant.

A particularly remarkable quality of this space is the level of interaction afforded to—indeed, required of—performance techniques: Once a sound is introduced, it will remain audible for a hauntingly—or embarrassingly—long time. This feature allows chords to be easily built by single-voiced instruments, and harmonies to be wildly improvised. On the other hand, if a mistake was made, it, too, would be mightily apparent for all to hear.

---

## Example 4

Classical Guitar performed by Steven Novacek in St. Stephen's Parish church, Seattle, WA. Recorded with a pair of Neumann KM 130 microphones, with about two meter equilateral spacing.

*from Ambassador ARC 1007*

**Discussion** Although most of its acoustic energy is output from the rose, a guitar is not a point-source instrument, as its soundboard propagates a complex wavefront (See Figure 2.). Nor is it, strictly speaking, asymmetrical: Performance requirements often distinguish between the right hand and the left. For the most part, however, these characteristics are apparent only to the performer; for most listeners in an enjoyable acoustic environment, any directional features are swamped by spatial qualities. For pleasing envelopment, lateral reflections are strongly favored.

In the present example, the space had greater than a three second reverberation time at mid frequencies, but had a warm sound associated with a substantial amount of IACC. But at much distance, clarity would have been a problem. The capsules were far enough apart that IASDs of the arriving indirect sounds were appreciable, even without the shadowing afforded by binaural techniques. Because of this, and the equally considerable IATDs and IAADs, the space pair resulted in a very spacious quality, but, without the "hole-in-the-middle" often associated with a such a technique, as the direct sound was very similar at the two microphones.

## Example 5

Trombone, didjeridu, accordion, garden hose, metal pipes, and voices in a large underground cistern at Fort Warden, Port Townsend, WA. Performed by Stuart Dempster, Pauline Oliveros, and Panaiotis. Recorded with four Brüel and Kjær 4006 omnidirectional and two Schoeps CMT-35 cardioid microphones. A later session was recorded by three independent systems, including one binaural one engineered by Gordon Hempton.

*from New Albion NA 022*

*CD*

**Discussion** Past a certain level of diffusion, "envelopment" takes on a new—in this case, almost other-worldly—meaning.

The objective qualities of the cistern location are, while unusual, nevertheless straightforward enough: It is a concrete cylinder which in an earlier incarnation held about two million gallons of water, and is 186 feet in diameter, 14 feet in height, has 89 supporting pillars—and it has a mid-frequency reverberation time of about 45 seconds (See Figure 3.), when damp (more when dry). In this environment the direct sounds represent a small fraction of the total acoustic energy arriving at the ears of

---

are typically recorded monophonically, given, in mixing, pan-pot treatment to develop a sound stage appropriate to the visual perspective and the sound designer's whim.

It is the author's position that, within practical limitations, all SFX should be recorded at least stereophonically—if not quadrophonically, allowing even more post-production flexibility (especially considering that the eventual utilization of the effect is not always known at the time of its recording). Coincident configurations are mono-compatible (which, while an over-rated commodity in audio-only presentations is often quite desirable when dealing with the vagaries of video and film post-production) and allow for a certain amount of spatial processing. Ambisonic and other surround sound techniques are sometimes indicated, and often preferred, although it should be noted that much SFX gathering is “guerrilla audio”; hence logistic considerations often render complicated schemes less useful: Single-point stereo microphones, plugged directly into a portable recorder, have the advantage of being very portable, allowing for (single) hand-held operation. This is of substantial practical import for the recordist hanging from the side of a precipice or wading through violent surfs.

While much surf sound is, like thunder, basically non-directional, many associated components are more wide-band. These include splashes against rocks, the water receding past small rocks, foam action, and, in some areas, wind in the nearby vegetation. These elements have, in natural audition, substantial amounts of IAADs, IATDs, and IASDs, which information coincident microphones do not, except for the amplitude differences, retain. In addition, directional pattern microphones usually are deficient in their very low frequency sensitivities. For realistic presentation, therefore, they require much processing if recorded with coincident microphones.<sup>16</sup>

Complex sound designs call for a layering of SFX along with dialog, Foley, and music. Thus, multiple aural perspectives can be accommodated, and desired amounts of simulated spatial cues can be introduced in independent layers. In the present example, several methods were used. First, because the original recording was done in an M-S configuration, it was a simple matter to boost the low end in the difference channel. Then, overall spatial equalization was performed (with a Lexicon 300 processor; see [11] for a discussion of this process). Finally, layering was employed, adding, among other things, the effect of a very large bass drum recorded with widely spaced Neumann KM 130 microphones. This added a good deal of out-of-phase low frequency signal, simulating different onset times for the sound envelope of the booming surf.

---

16. Directional imaging from coincident microphones can be quite accurate, as they are sensitive to amplitude differences. However, they retain very little information about spectral differences, and none at all, in theory, about time differences. Therefore, spatial imaging is not possible using them.

---

qualities depend, in turn, on the characteristics and position in space of the lightning producing them. Thus, embedded in this signature are cues for distance and intensity, but not, for the most part, direction. (The latter feature can be determined by IAADs and IATDs, or relative phase, along with interaural spectral differences [IASDs]. However, these differences are unreliable or non-existent at low frequencies, and for all but the nearest thunder, the spectral signature is largely devoid of content above 60-70 Hz.)

The storm's power, on the other hand, can be inferred from the thunder's frequency (the lower, the mightier) and the duration of its roll (the longer the roll, the longer the lightning channel). Proximity can be determined, in the absence of visual cues, by noting spectral information. Distant thunder contains fewer harmonics, due not only to the attenuation of high frequencies because of air resistance, but also to acoustic refraction which carries the upper frequencies—more than lower—up and out of earshot.

Thunder which is very near does, in fact, contain considerable high frequency information, due to 1) the impulsive nature of the advancing wavefront, 2) less attenuation and refraction, and 3) envelope pulsations caused by the superimpositions of each of the lightning channel's segments. (See Figure 1.) These cues can indicate direction, but do so only at a time when the only direction of the immediate concern to the listener is that toward the nearest safe shelter.

Scientific studies of thunder signatures utilize an array of widely spaced microphones to re-create the lightning channel in three dimensions [43]. As this methodology is unavailable to the unaided organism with one ear on either side of an ordinary human-sized cranium, such spatialization must be considered unnecessary in an evolutionary sense. Thus, one might be tempted to record thunder monophonically. However, in addition to proximity and intensity, the human listener is also interested in context: Do the birds still chirp? How hard is the rain? Is there anything burning? For this reason, spatial sensitivity is desirable; thus, spaced, binaural, and quasi-binaural configurations, all of which retain a good deal of spatial qualities while maintaining a full bandwidth are recommended, precluding the use of directional microphones.

### Example 3

**Ocean surf against rocks, on the Pacific Coast, south of Cape Flattery, Washington. Recorded with Sanken CMS-7 stereo microphone in the M-S configuration at 110°.**

*for a Nike multi-media presentation*

**Discussion** Film, video, and multi-media presentations have, by definition, a good deal of visual information; hence, spatial cues are less important than for audio-only presentations. Further, the creative design often calls for more post-production flexibility than production accuracy. Therefore, most sound effects (SFX) for them

---

## Example 1

Forest ambience at dusk in Mount Baker/Snoqualmie National Forest, recorded by Jonathan Storm with a near-coincident pair of Sennheiser MKH 40 microphones. We hear a nesting pair of spotted owls prepare for their nightly activities. Various other birds sing above the roaring river.

*from EARTH TUNES ET5-1004*

**Discussion** A forest ambience is an all-enveloping one, in four-dimensions including a very important temporal one. Things crawl. Things fly. Things run, fall, hunt, splash, mate, eat—all according to the demands of the seasons. Of all Earth's fauna, we, as humans, are the only ones who substantially alter the environment, forcing it into conformity with our desires. All other creatures must take the environment largely as a given, and come to terms with it. It is only natural, then, that a forest soundscape should be represented with as many realistic spatial cues as possible. This calls for an extreme dynamic range, off-axis tonal shadings, a wide spectral range, and a reduced but critical rear sensitivity ("rear", that is, relative to the listening axis, not the microphone capsule). For loudspeaker presentation, near-coincident techniques satisfy all these requirements fairly well (their weakest feature, in this regard, is that off-axis colorations vary considerably among specific transducers) although their directional microphones will exhibit low frequency roll-off, limiting their performance in some environments. Binaural and quasi-binaural techniques work quite well, assuming the use of high quality microphones. The latter configurations often have the advantage of an enhanced low-end, plus the practical benefits of lower handling and wind noise than do discrete directional microphones.

A further interesting note about this example is that it had, in the event, forensic value: In the political squabbling over which forest lands to preserve for the habitat of the spotted owl, this recording established the presence of a nesting pair at a specific site.

## Example 2

Spring afternoon rolling thunder. Recorded in Eastern Kansas by Gordon Hempton with "Fritz", a Neumann KU-81 artificial head.

*from Peter Roberts Productions PRCD 2008*

**Discussion** Apart from the rare cataclysmic geological event and some rock concerts, thunderstorms produce the most intense sound Earth's creatures are likely to experience. To humans, thunder can range from the benign to the truly frightening, depending on, among other things, how near and how powerful the storm is. The signature of the thunder peal depends on its frequency, its amplitude, and the temporal ordering of the waves perceived at any given observation site. These

---

of survival [35, 36, 37, 38], and gives rise to commercial activities (including, of course, the business of audio) of vast scope and scale.

To reprise, in this context, the question of why we prefer some sounds over others, we can say that survival, art, and commerce are, like the corporeal senses, interrelated and interactive. A few physical correlates of “beauty” are known. For instance, we know that listeners do not enjoy listening to sounds in an anechoic environment. They prefer a reverberant field—and, especially, one which has a substantial amount of lateral reflections [3, 4].<sup>14</sup>

Objective research along these lines has recently intensified, especially given the impetus of impending adoptions of High Definition Television standards [39]. But many questions remain, and much is yet to be studied and documented.

## Examples

The Abstract of this paper notwithstanding, a valid position can be taken that, while some audiographic materials may be more in vogue than others at any particular moment in history, there are, in fact, no “non-standard” recording situations: Anything that can be recorded is allowed as standard practice.<sup>15</sup> Nevertheless, production aesthetics are not universal, and one engineer’s “mainstream” is another’s “underground”. Thus, a kind of “cross-cultural” audio exegesis can be informative. The following examples are presented in an exposition/discussion format, and cover the following categories of “stage-less” audio: Environmental sounds recorded for audio-only playback; environmental sounds recorded as multi-media sound effects; solo music in a diffuse field; and ensemble music in an extremely diffuse field.

(It is assumed that the reader is familiar with the standard techniques of stereophonic and surround recording involving coincident, near-coincident, spaced, binaural, and ambisonic configurations. For a well-annotated synopsis of them, see the paper of Streicher and Dooley [40]. Also valuable are the references of White [41] and Eargle [2, 42].)

---

14. Lateral reflections, particularly at angles of around 60° off axis, tend to maximize interaural cross correlations (IACCs), from which we can infer spatial information. Is it that we sense potential danger in an environment which does not fully identify itself? It is sometimes alleged that we, as humans, are at the top of our food chain, but is this perhaps only correct in our evolutionary “comfort zone”? Or, is there a more metaphorical “food chain” involved, a cannibalistic one of the dog-eat-dog school of human “turf-battle” interaction? (And, are these questions art or are they science?)

15. And it is not only allowed, but essentially required, if there is a paying customer.

---

to similar ones in the organism's library of experiences, it can thus determine both the nature of these acoustic events and their closeness.

A further acoustic cue relating to proximity involves the perception of the amount of diffusion in the soundfield. Confined environments, such as caves, are reverberant in varying degrees according to the composition of their boundaries, while prairies are almost totally free-field. The signature of a diffuse field sound can tell the listener much about its size, and the proximity and character of walls and ceilings, and openings.

Finally, localization becomes consequential. A predator must know the exact location of its prey, or, more specifically, where the prey will be upon closure. But, in this, hearing is usually only an auxiliary sense. More important are vision and experience (as in knowing which way to run for optimal results).<sup>12</sup>

Of course, hearing can be used independently, if necessary: For example, a nocturnal raptor, such as an owl, ordinarily relies heavily on exceptional visual acuity and sensitivity. However, in the absence of light, it can still successfully hunt on acoustic cues alone [33], although with a somewhat degraded performance. Even when hearing is the primary mode for localization, a passive, fixed-head technique (i.e., the conceptual basis for stereophonic presentation) is never employed, except by acousticians, headphone wearers, royal guards, and recording engineers. Bats echolocate. Humans turn their heads, thereby putting different elements of the soundfield in the interaural shadow, while looking toward the direct sound of interest. [29]<sup>13</sup>

For evolution, transcendence is a luxury, for the objective researcher, a bane. Still, "truth" (or "beauty") is important for humans, at least,<sup>14</sup> a concept which Diane Ackerman [31] goes on to elucidate convincingly. This urge to create, critique, and consume abstract and illusory, albeit purposeful, webs of sensory stimulation is universal among all cultures and societies, even those living closest to the margins

---

12. Experience and adaptability are also compensatory "senses" in that they "even the playing field" for those organisms whose physical abilities are in decline. As the saying goes, "old age and treachery will overcome youth and skill."

13. Hardison states, "...a recording reduces the listener to the position of passive consumer." [28] But he is wrong: Even with "passive" listening (and most concert-goers are no more "active" than the stay-at-home stereo buff), we retain a large amount of interactive control over the process: If we find the music not to our liking, we can, depending on the nature of the offending source material, turn in our season tickets, bribe the Music Director, hire a louder band, pull the plug, take a hike, or call the police.

14. It can perhaps be argued that an artistic sense is merely a complication of the survival instinct. And, in fact, so-called primitive art does seem to be closely associated with fertility, a good hunt, and other matters bearing on existence in the trenches of life. Is it art or is it religion? Regardless of the semantics of the issue, we are not demonstrably alone among the Earth's fauna in using sound as more than a tool of raw subsistence. Wolves, for example, have a complex vocalizing ability, used in a wide variety of social activities [34].)

---

“...But presently the gray dawn stole over the world, the birds piped up, then the sun rose and poured light and comfort all around, everything was fresh and dewey and fragrant, and life was a boon again.”

—*The Autobiography of Mark Twain*

“There was a song in every heart; and if the heart was young, the music issued at the lips.”

—*The Adventures of Tom Sawyer*

## The Soundscape Dissected

At this point, it becomes necessary to be partially reductionist, the above disclaimer notwithstanding. Ignoring aesthetics, we may ask what the important elements composing the soundscape are.<sup>10</sup>

At the root level is content, the primacy of which is evidenced in several everyday observations. For example, we tolerate monophonic radio broadcasts and sound reinforcement systems<sup>11</sup> easily, especially with the spoken word. More fundamentally, does, for the listener, the sound suggest having lunch or being lunch? Is there danger ahead, or prosperity? Is the man discussing a woman’s smooth skin and moist lips referring to his own wife, his neighbor’s, or the listener’s?

These substantial questions have a partial, if not always sufficient or accurate, answer from a real-time analysis of the soundscape. Much of the relevant cues are spectral: Low frequencies typically represent large (frequently predatory) animals. Lions, for instance, roar (and purr); mice do not. And extremely low frequencies can signify geologic events of note, such as earthquakes or volcanic eruptions. Stress, possibly indicating, in humans, the intent to deceive the listener, skews a speaker’s vocal harmonic composition.

Content also has different meanings according to temporal contexts. The rains of April do not sound the same as the rains of November, and these differences are only partially due to different intensities, frequencies, and rhythms. On a more human level, we can muse that common environmental sounds, such as dogs barking, water dripping, and doors creaking then shutting softly, mean something much different in the dead of night than they do during the day.

The next level of importance involves proximity. First an organism needs to know if what is sharing its space is friend or foe, benign or dangerous, tasty or toxic. Then, how close is it? Again, spectral cues are important: Since high frequencies attenuate more quickly over distance than lows, the nearer events are audibly brighter than the distant ones. By cross-referencing the envelope of the sound event

---

10. Stebbins’ [32] discussion of these features for non-human fauna can be used as a model for exploring human utilization.

11. ...As well as, to our occasional chagrin, background, or elevator, music....

---

subject moves to a still closer location, the head-related transfer function (HRTF) becomes increasingly significant; changes in the frequency spectrum suggest a previously undetected depth dimension, permitting spatialization. Finally, from within the orchestra, three complete spatial dimensions are discerned. (Real acoustic situations are intrinsically complicated by the fact that they never occur in truly free fields; to a greater or lesser degree indirect sounds are always present. The natural habitat of an orchestra is, of course, a large auditorium. In the present model, however, this merely means that dimensionality is sensed earlier, proportionately to the amount of indirect sound present.)

Thus it can be seen that, considering only the spatial aspects, a performing musician will have a far different auditory “view” from that of either the conductor or a typical member of the audience. Which is “correct” depends entirely on who does the defining. Again, that person is usually the one who is paying the bills.<sup>8</sup>

However, sound, as used in real life, consists of far more than spatial cues. As a philosophical point of departure, we could, perhaps, reflect that using the auditory sense to provide aesthetic pleasure must be a relatively new development in our evolution. That utilization is artifactual, in the sense that it is unnecessary for basic subsistence.<sup>9</sup> Essentially, “the body’s quest isn’t for truth, it’s for survival.” [31]

If we posit at the heart of the matter a survival purpose for hearing, then it seems reasonable to consider all events, spatial, temporal, and contextual, within the listener’s soundfield to be potentially important, and in an integrated way. We find evidence of this in literature. Mark Twain, an astute observer of the role of the senses, wrote

“Profound silence; silence so deep that even their breathings were conspicuous in the hush. Tom shouted. The call went echoing down the empty aisles, and died out in the distance in a faint sound that resembled a ripple of mocking laughter.” “...The blackness of darkness reigned, the perfect stillness was interrupted only by the occasional mutterings of distant thunder.”

—*The Adventures of Tom Sawyer*

therein setting a remarkable tone of terror-inspiring desolation. Survival favors the organism with a finely-tuned reaction to fear and the unknown. But the flip side of this equation, a sensitivity to the prospects for prosperity, also obtains:

---

8. It is tempting to speculate on how the audio industry would have progressed if its early popularizers, Edison and Berliner, had been musicians, since the latter have a different acoustic perspective than do audience members.

9. It is in a similar vein that we can view the approach of the epicurean to food consumption, or of the romantic to sexual intercourse—which is to note that what is “raw survival” to some is “life not worth living” to others.)

---

transduced via bone conduction. For the musician, music is very much tactile. and proprioceptive)

- “In a concert, audiences can influence the performers by hissing or applauding or coughing or throwing roses onstage.” [28] Also, they can give—or withhold—money.
- In contemporary rock concerts, the audience members feel cheated if they do not leave with, not only their ears, but also their skin, tingling.
- The body ascertains sound direction not only by the classically researched mechanisms of inter-aural differences of time, amplitude, and spectral components, but also from a host of visual, tactile, vestibular, and proprioceptive cues. Further, the same thing works in reverse: Auditory stimulation influences the body’s apparent orientation [29].
- Echolocation is a well-known phenomenon of the hunting practices of bats, nighthawks, dolphins, attack submarines, and other low-light predators. Dogs bark; toddlers cry, whoop, and yell; recording engineers employ hand-clapping or shouting to determine the characteristics of an acoustic space; “Mary” might simply say, “Hey, Joe, where are you?”

Microphone technology has a difficult time accurately documenting such soundscape phenomena. Thus, for the practical recordist it is necessary to adopt a model employing the supposed, or paradigmatically correct,<sup>6</sup> salient cues and rejecting those seen as non-essential. Note that in doing so, we have already made critical decisions which reflect our artistic, intellectual, and commercial agendas, but are otherwise arbitrary. Who says localization is more important than spatial qualities? The engineer—or the producer, or the reviewer, or, ultimately, the client—that’s who. For that matter, who says the sound of interest is in the frontal stage, oriented along a basically one dimensional line, with an occasional foray into one or two additional dimensions? The answer is the same.

This is to say, the answer, leaving out commercial aspects, will be different, even disregarding non-auditory cues, for different observers. For example, a symphony orchestra is ordinarily thought of as a non point-source acoustic stimulus. But consider its spatial dimensionality as seen by a hypothetical listener who is allowed to move about (a normal hearing ability is assumed).<sup>7</sup> In a free field, a distant observer will consider the performance to be emanating from a single point in space. At a nearer perspective, this point will spread out, left and right, as interaural time and amplitude differences (IATDs and IATDs) begin to allow localization. If our

---

6. It may be noted that “P.C.”, paradigmatically correct, can also designate the expression, “politically correct”. It may further be noted that the coincidence is not always devoid of ironic merit.

7. This is an analogous thought experiment to the one suggested by Mandelbrot [30] to illustrate the concept of fractional spatial dimensions.

---

of physical phenomena, such as audio, cannot thrive.

Nevertheless, standard procedures can betray, however unintentionally, the human tendency to map results onto a preferred theory. This is not the same as saying that researchers regularly skew their findings; simply, a hypothesis must be chosen, and, naturally, it must reflect the philosophical stance of its chooser.

Particularly elegant paradigms permit fairly accurate extrapolations. But testable hypotheses require, by definition, reductionist analyses. It is suggested here that “real-life” sensory phenomena are inter-modal, synergistic, and, most importantly, interactive, and hence do not willingly give themselves over to reductionism. Moreover, theory-referred analyses sometimes lead, it may be bluntly said, to counter-productive results. This is especially true when the theorist has a good deal of intellectual stature.<sup>5</sup>

To reiterate, the point is not to condemn traditional methodology, but to suggest ways that it be incorporated, through the employment of anecdotes and descriptions, into the fuzzy gestalt (to extend Thiele’s description [26]) of everyday phenomena).

## Soundfields and their Perceptual Correlates, Soundscapes

In an attempt to record a soundfield, the first challenge is to define it, to choose the relevant aspects. This is not a trivial undertaking, as the totality of any soundfield in context (i.e., the soundscape) is unwieldy in the extreme, being, as noted earlier, intermodal, synergistic, and interactive. To illustrate:

- The “exit sign effect”, noted by sound designer Ben Burt [27], wherein audio that tracks visual action laterally across the screen seems to continue following the action after it leaves the screen.
- An instrumentalist, like a harpist or ‘cellist, will often complain that “no one records my instrument the way it really sounds.” (Indeed, no one can, because much of the characteristic sound enamored of this musician is

---

<sup>5</sup> An appropriate example of this involves the work of Stanley Lipshitz. In the mind of the author, Lipshitz’ credo, “I believe that spaced-microphone recording techniques are fundamentally flawed, although highly regarded in some quarters, and that coincident-microphone recordings are the way to go [23],” went against all intuition and experience: For some reason, almost all customers, with the exception of the occasional broadcast chief engineer, strongly prefer the results produced by a non-coincident technique. But it remained for others to show where the theory did not hold. (Principally, Lipshitz misapplied the principal of summing localization [24]. Also, he ignored the fact that head-referred audition is uncorrelated above 700 Hz. (lower, if using more realistic spherical or elliptical head models—and, of importance to this paper, as low as 200 Hz.) in loudspeaker presentation for images outside the loudspeaker space. As it turns out, frequency-independent models of localization cannot be produced [24]. Ironically, this can be inferred from a close reading of Blumlein’s classic patent [25], to which Lipshitz’ work can claim a certain lineage.) But in the meantime his analysis was widely propagated, and is still cited by many proponents of coincident microphone techniques.

---

soundfield is not significant; the listener could not tell a “mirror image” from the original. Or the dominant signal may be indirect and in a diffuse field (see Example 5). Other “stageless” situations include environmental recording (see Examples 1 and 2), sound effects work (see Example 3) and most popular recording<sup>4</sup>, ethnomusicological studies, and much acoustic measurement. With the exception (in most cases) of the last two categories, what we are particularly interested in for the original recording is the soundfield, where left and right in the forward direction may play only a minor role, or none at all.

Interestingly, many techniques evolved for fixed frontal stages are, or can be coerced into being, appropriate for soundfield applications as well. In the Examples section, some of these procedures will be discussed in the context of actual recordings made for genuine paying customers.

## Definitions

Before proceeding, it may be beneficial to define two terms frequently used in this paper.

SOUNDFIELD refers to the physical aural cues available to an organism at some point in time and space. The elements of a soundfield can be objectively measured and manipulated, but are not dependent on the actual perception of the organism.

For present purposes SOUNDSCAPE is a superset of SOUNDFIELD, and includes perceptual aspects, which in turn include cues from other corporeal modes (vision, olfaction, etc.), context, experience, and whatever else might influence perception. The soundscape of an organism is, therefore, much more difficult to pin down quantitatively, and yet, of critical importance to its (the organism’s) well-being.

## Methodology

This paper presents no new “findings”, as such—no results of paradigmatic research procedures in the academic sense. Rather, the intent is to “stir up the pot” of existing knowledge by way of providing alternative views of the soundscape as interpreted by the listener. Thus, a few comments on methodology are in order.

The physical anthropologist Harry Shapiro states that “rigorous analytical methods guarantee some measure of the objectivity without which the less weary observer is often led into sentiment and bias.” [22] At the same time, Shapiro contends that, where human complexities are concerned, anecdotal and descriptive methods can teach us at least as much. The present author has no quarrel with either approach. Without a rigorously objective approach, a discipline based on the study

---

4. In both sound effects and popular recording, a stage may, in fact, be defined (by the use of pan-pot placement in the mix), but not during the original recording process.

---

In any case, most listeners derive more pleasure from an aural experience if some sense of space is presented [3, 4]. Snow [5] commented on depth considerations, although, curiously, he evidently based his conclusions on a single ear model of hearing, sensitive to only quality and intensity. And in recent years spatial aspects of a recording have been the subject of much commendable research.<sup>2</sup> Some investigations have resulted in the development of commercially useful products [7, 8, 9], while many others have concentrated on processing schemes, often in an attempt to transmute the acknowledged virtues of binaural recording techniques to the stereophonic listening environment [10, 11, 12, 13, 14, 15]. Still others [16, 17, 18] do both. Finally, there have been some suggestions regarding the possible viability of a three dimensional presentation [19, 20].

Also, computer modeling has resulted in major advances in the understanding of human localization and spatialization [21].

Thus, a review of the literature suggests strongly that both the technology and art of soundfield recording, which suggests that the listener is sensitive to aural cues from any direction, have come a long way.

Still, while given increasing due in the literature, spatial qualities nevertheless seem usually to be considered as merely features which add ambience, or “color” to the direct sound. The concept of a stage, either an actual one in the case of unmixed acoustical music or a virtual one designed in the studio, seems more or less universal in the audio business, and, for the most part, unarguable.<sup>3</sup> Certainly, the idea of being able to localize a sound image has applicability when the position of that image on the stage is aesthetically important. For example, in symphonic recording it is rather disconcerting, so to speak, to hear a solo violin from the far right during stereo playback. Further, even when the stage is less historically defined, as might be the case with some non-orchestral ensembles or certain vocal music, a certain amount of stylistic consistency is usually deemed desirable. Shifting a specific instrument or voice arbitrarily across the stage within a song (or even between songs), is, in most cases, unacceptably jarring.

But there is a wide range of recordable sounds for which the presumption of a stage, at least in a left-right sense, has little or no bearing. For example, the voice, and many musical instruments, such as the harp and most wind instruments, are acoustically symmetrical about the vertical axis. For their solo repertoire, left and right are therefore meaningless (see Example 4). Even when ensembles with an obvious asymmetry are recorded, the specific placement of each element within the

---

2. Blauert’s now classic text [6] could be thought of as patriarchal in this regard.

3. One can perceive this in the use of the terms HEADPHONE PRESENTATION and LOUDSPEAKER PRESENTATION: Musicians, for instance, present (with the engineer, producer, et. al., as mediators); the consumers buy into it. The concept is also implicit in the oft-stated goal of live recording: to simulate the sound occurring at the “ideal” (if non-existent) listening location relative to the “soundstage”.

# Towards an Understanding of Soundfields and The Development of their Audio Representations

*Albert G. Swanson  
Seattle, Washington USA*

Most techniques of stereo recording have developed for fixed stage applications. That is, they are designed to simulate a listener's relationship to an audio event, preserving as much spatial information—depth, left-right imaging, and room characteristics—as is deemed desirable. Much debate is made of the technical elegance and aesthetic appropriateness of one technique or another. For many real-life applications, however, these are irrelevant. This study explores some such situations, and discusses how standard stereo procedures have been used to simulate soundfields. Finally, proposals are appended concerning the application of quadrasonic techniques to fully represent soundfields in two and three dimensions.

## Introduction

IMAGING; LOCALIZATION; “STEREO SPREAD”; DEPTH; and, more recently, SPATIAL RESOLUTION: With varying degrees of objective value, these terms have established themselves as major elements in the lexicon of audiology. From early in the history of stereophonic recording [1], much discussion has centered on which of these parameters are the most important aesthetically, and what techniques best satisfy the circumscribed objectives.

Many quite elegant procedures have thus evolved (See Eargle's reference [2]). But as each is associated with both a cadre of supporters and an approximately equal number of detractors, the various debates are far from decided. Apart from the technical arguments are the philosophical ones: Should, for instance, the engineer at all costs maintain the integrity of the stereo image, or is a little blurring acceptable, in order to introduce a sense of “spaciousness”? For all such questions the answer is: It all depends—on not only engineering aims, but on more mundane criteria: Who is paying for the product? Who is selling it? And, sometimes most importantly, who is reviewing it?<sup>1</sup>

---

1. This was brought home to the author early in his career, a time when most of the clients were the performing musicians themselves. Since the customers pay the bills, there naturally is a strong incentive to heed their opinions. But these “customers” had not, for the most part, read much of the technical literature. Hence, they typically were unaware of the demonstrated advantages and proven virtues of one technique over another. The only criteria available to them were subjective. Later, of course, other agendas had to be accommodated. Naturally, with experience comes reputation and stature, and hence authority, by virtue of which artistic decisions can be mandated. But the point remains the same: Right or wrong, “the customer is always the customer.”