WHY THINGS DON'T WORK: WHAT YOU NEED TO KNOW ABOUT SPATIAL AUDIO

Gary S. Kendall
Sonic Arts Research Center
Queen's University Belfast
garyskendall@me.com

Andrés Cabrera
Sonic Arts Research Center
Queen's University Belfast
mantaraya36@gmail.com

ABSTRACT

Composers working in the sonic arts have frequently found themselves attempting to use spatial audio in ways that didn't work as intended. Maybe more than any other facet of technological music, mastering spatial audio seems to involve a learning process in which one slowly discovers the things that work and those that don't. The purpose of this paper is to foster understanding of spatial audio through examples of practical problems. These problems include both some general misconceptions about spatial hearing and some specific examples of things gone wrong. A particular lesson to be learned from this discussion is that there is no silver bullet for solving spatial audio problems, and every situation needs to be understood in appropriate terms.

1. INTRODUCTION

Composers working in the sonic arts have frequently found themselves attempting to use spatial audio in ways that didn't work as intended. Maybe more than any other facet of technological music, mastering spatial audio seems to involve a learning process in which one slowly discovers the things that work and those that don't. That this learning process is so tentative and empirical reflects the lack of a conceptual foundation that could guide the artist when conceiving spatial ideas and when translating these ideas into practice. It is also easy to be misled by preconceptions about how spatial audio should work. The purpose of this paper is to approach conceptual understanding through practical problems.

The early pioneers of electroacoustic music pushed the frontiers of spatial audio and achieved monumental successes in the artistic use of space. Varese, Stockhausen, Schaeffer and Poullin, Bayle with the Acousmonium, Chowning and onwards---spatial audio has been an expanding area of artistic expression. On the other hand, the great advantages in computer and audio technology that we enjoy today have not necessarily led to greater advances in spatial audio. Quite possibly, pushing back the frontiers of spatial audio today depends more on understanding spatial perception and cognition than on raw computing power.

2. GENERAL MISCONCEPTIONS

In everyday life, every person is able to navigate around in a spatial world, to talk about space and even to imagine unknown spaces; spatial thinking is one of our most deeply embedded cognitive capacities. The ease with which we think about space is possibly a miscue to how easily spatial ideas can be translated into spatial audio, which has its own unique capacities, intrinsic nature and inherent limitations. While there are many ways in which hearing participates in our everyday spatial thinking, sonic artists need to be especially alert to the nuances and idiosyncratic role of spatial hearing in spatial thinking. Not every spatial idea can be reverse engineered into sound.

Clearly, our expectations about spatial audio should be in alignment with the fundamental capacities of the spatial hearing. For example, consider how auditory spatial acuity varies with the direction of the sound source. In front of the listener horizontal *localization blur* is ±3.6°, and to the sides ±9-10°. Above the head and slightly to the rear, vertical localization blur is ±22° [3]. Apparently, what listeners perceive is not well described as a point source. More appropriately listeners' perceptions can be described in terms of a small set of auditory spatial attributes. Following work by Rumsey [16], Kendall [11] offered the image in Figure 1 as an illustration of these attributes.

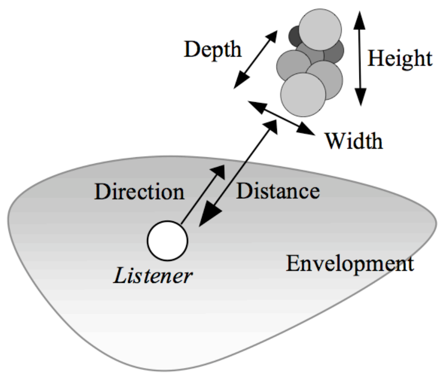

Figure 1. Auditory spatial attributes (from [11]).

Important for spatial audio is the primary role of envelopment as a spatial attribute. Envelopment has been the focus much research in spatial audio [5, 20]. A special caveat to offer is the observation that there is no clear separation between auditory width and envelopment; one can blend into the other.

3. WHAT WENT WRONG?

The way that most composers and audio engineers discover their misconceptions about spatial audio is through direct encounters with things that don't work.

3.1. Why doesn't the sound image get broader when I distribute a signal to three of more adjacent loudspeakers?

Everybody learns about the precedence effect, but the practical consequences of it often take time to absorb. The precedence effect is a psychoacoustic phenomenon in which the auditory system gives precedence to the first arriving sound [2, 7, 19]. It is usually described as an aid to spatial perception in reverberant spaces, where the direct sound from a sound source reaches the listener before the reflected sound. The listener's spatial image localization is dominated by the first-arriving/direct sound, and the perception of the early reflected sound is suppressed.

The precedence effect works the same with multiple loudspeakers: the listener's spatial image is determined by the first-arriving sound from one loudspeaker and the later arriving sound from the loudspeakers is largely suppressed. When a source signal is distributed among multiple loudspeakers, a listener will hear the source as emanating from the closest loudspeaker. (See [13] for an excellent review of the precedence effect.) The primary way to overcome this limitation is to trick the auditory system into believing that the delayed sounds are not reflections. This can be done by decorrelating the signals sent to each loudspeaker [9, 10].

3.2 Why does my dynamic panning work sometimes and not others?

When a listener is equidistant from all loudspeakers, in the location called the *sweet spot*, there is no one loudspeaker from which sound arrives first, and therefore, precedence has no effect. Spatial panning systems, whether amplitude panning, power panning, Ambisonics or VBAP, are all affected by the precedence effect when listeners are closer to one loudspeaker than another. In all multichannel-reproduction settings, that includes everybody except the person in the sweet spot.

Because of the precedence effect, all these listeners will experience a discontinuity instead of a smooth panning as the amplitude in the closer loudspeakers is enough to trigger the precedence effect. The degree of the discontinuity depends on the exact distances of the loudspeakers from the listener because these distances determine the time delay of the sounds reaching the listener. The degree to which a continuous spatial path is perceived is particularly dependent on the transient content of the sound material. Sound sources with rapid re-attacks and high-frequency content are generally the most successful. Then too, plausibility and comprehensibility affect whether the listener understands the changing sound as spatial motion along a path.

3.3. Why did the spatialization in my piece sound so different in a large space than it did in the studio?

Although differences in room acoustics (reverberation and coloration) can cause some differences, most major changes in perceived spatialization are due again to the precedence effect. An important property of precedence is that it ceases to affect perception after a certain time delay. Depending on the nature of the sound material, that delay may vary anywhere from 5 to 50 msec [13]. When one monitors inside of a small studio, the differences in the distances of the loudspeakers generally creates delays of less than 5 msec. In a large space, the delays can easily range up to 12 msec or more. The change in the length of the delays affects which sounds are affected by precedence. This is true even when the relative angles of the loudspeakers are exactly matching. In a large space the delays are different for nearly every listener. Another important factor to bear in mind is that it is easier to be located in the sweet spot in a smaller environment than a large one.

3.4. Why doesn't my circular panning work?

Composers working with multiple loudspeakers frequently aspire to move the position of sound sources around the listener by use of multichannel panning. Nearly everyone has acquired their understanding of panning from working in stereo with one loudspeaker to their left and another to their right. With this kind of stereo, one experiences a *phantom image* that can be positioned between the two loudspeakers. (This is approximately the same whether the loudspeakers are in front or in back of the listener.) However, that experience does not transfer to situations in which there are two loudspeakers to one side of the head. It is not that sounds cannot be positioned on the side, but rather that there is seldom a coherent phantom image: the image is spread in what Kendall [11] describes as *image dispersion*. This dispersion blurs the location of the sound image in a way that depends on the transient and spectral characteristics of the source. This becomes particularly evident with broadband sources that cause different parts of the source sound to be biased toward the front or the rear. This is the reason that Tom Holman [8] recommends against panning between the front side and surround loudspeakers in 5.1 systems. This effect is also difficult to predict and will depend on the precise location of the listener.

It is easy to understand how image dispersion makes it difficult to create an effective impression of circular motion around the listener with arbitrary sound sources. As mentioned before, some source characteristics are more favourable to perceived motion than others (narrow-band, high-frequency, transients). Cognitive factors also help if the listener is able to apprehend an apparent path to the motion. On the other hand, it should also be mentioned that image dispersion has long been used in live diffusion to create a broad spatial effect. When a signal is placed in both the front and rear loudspeakers, image dispersion creates the impression of sound enveloping the audience [1].

3.5. Why don't the recordings I make with binaural mics sound the same as being there?

The stereo signals recorded with most portable or clip-on binaural mics capture the separation of the ears and will include *interaural differences* in intensity and time. Because these mics also capture the acoustics of the head, the interarual differences will also be frequency dependent. For example, the head affects interaural intensity differences because it is a more effective block for high frequencies than for low frequencies. It affects interaural time differences too, but in a way that is too complex to describe here.

When you listen on headphones to recordings made this way, you get a pretty good idea of how these acoustic factors affect spatial perception. You will probably hear sound images to the far sides, behind you and maybe above you, but not very often in front of you. That is the main problem with binaural recordings: something about the experience is quite different from being there. That difference is largely caused by the lack of coordination between body movement and acoustic changes at the ears. When the head moves, the sound at the ears should change in an appropriate way. Without ear acoustics being coordinated with head movement, it is difficult to perceive sound sources directly in front of you. What typically happens is a kind of localization error in which frontal sounds are heard to the rear; this is called *front-back reversal*. It is also possible that sound images are stuck inside the head and there is a lack of *externalization*.

If you do the recording with a dummy head, or with special mics inserted into the ear canals, your recordings include all of the acoustic information that people use in everyday life. The composite of all this information is called a *head-related transfer function* (HRTF). These include the acoustic effect of the outer ears or pinnae. The pinnae have a major role in perceiving elevation and they assist in separating front from back. You can download laboratory recordings of HRTFs and use them to create binaural sound through convolution [12]. If you combine dynamic HRTF convolution with dynamic tracking of head position, frontal imagery is hugely improved. In fact, if you want to create a real sense of being there, render multiple sources in a virtual environment with HRTF convolution and head-tracking as is done with real-time simulators.

3.6. What goes wrong with the 3D effect when I play my binaural recordings over loudspeakers?

Binaural recordings or renderings have in them the HRTF information that encodes the direction of sound sources, but there is a big difference between listening with headphones and listening with stereo loudspeakers. When the signal for the left ear is reproduced by the left loudspeaker, two things go wrong. First, the signal arrives at the listener's left ear with the HRTF for the left loudspeaker location superimposed on the signal, effectively piling two HRTFs on top of each other and creating conflicting or inadequate cues! Second, the signal also crosses over to the right ear with that HRTF for the left loudspeaker and is added to the signal reaching the right ear from the right loudspeaker. This is the more catastrophic problem, the *crossover* of signals between the left and right sides.

It is not an easy problem to solve, but an excellent system for crosstalk cancellation was devised by Schroeder [17] back in 1963 and numerous modifications have been implemented since [6]. Many commercial systems also attempt to solve this problem. Without attempting some form of correction for crosstalk, the best you can hope for is that the sense of elevation in the original material comes through.

3.7. Why do phantom images seem higher than the loudspeakers?

Whenever listening to stereo loudspeakers, there are HRTFs for the loudspeaker directions and there is crosstalk. It doesn't matter whether you use amplitude panning, power panning or whatever: the effect of HRTFs and crosstalk is there. One result is that the elevation of the image is directly dependent on the source material. Spatial hearing cues to elevation are affected by the distribution of energy across frequency. High, bright sounds tend to localize higher than low, dark ones. This has been particularly well studied with filtered signals [4, 14, 15, 18]. That virtual images vary in their perceived elevation has long been known by recording engineers.

We should mention though that hearing virtual images of this sort depends on the acoustics of the reproduction environment. Spatial imagery can change dramatically especially depending on the presence of reflecting surfaces near the loudspeakers. Reflecting surfaces create additional signals reaching the ears and further ambiguating localization cues. A reproduction system can be particularly badly affected by environmental asymmetries in such reflections.

3.8. Why do I still hear single image when I put the harmonics of a sound in different loudspeakers?

This is an issue that comes up often with frequency domain processing where it is easy to break a sound source up into multiple frequency bands. Distributing those bands to multiple loudspeakers or dynamically panning the bands in independent trajectories usually does not create the intended effect. Imagine a situation in which a short performance on the cello is broken up into multiple frequency bands that are distributed among multiple loudspeakers. This is an example of a situation in which different categories of auditory cues are in conflict with one another. There is one set of cues that lead the auditory system to form a single, fused sound image of the cello and other cues that suggest multiple images in multiple locations. The fusion of the image wins and the listener most probably perceives the sound event as emanating from an indistinct, but nonetheless narrow, location.

In order to create a listening experience more akin to the intended effect, the fusion of the cello has to be broken down. This can be accomplished by de-synchronizing the partials, or adding contradictory vibrato patterns on the individual components. For the auditory system to recognize events in multiple locations, it first must recognize multiple events.

4. GETTING THINGS RIGHT

From the above discussion, we can observe that understanding a few key factors would help most composers to get their spatial audio right:

1. Precedence effect (Questions 3.1); its dependency on source characteristics (Question 3.2); its dependency on arrival-time difference (Question 3.3).
2. Idiosyncrasies of multi-loudspeaker localization (Question 3.4).
3. Directional Hearing and HRTFs: role of head movement (Question 3.5); understanding the role of HRTFs (Question 3.6 and 3.7)
4. Interrelationship of spatial hearing with auditory fusion and stream segregation (Question 3.8).

5. CONCLUSION

Experience is the master teacher and one hopes that its lessons lead us to an understanding of how to avoid future problems. The questions we have sought to answer here hopefully lead to an understanding of some underlying issues in employing spatial audio. A particular lesson to be learned from the above discussion is that there is no silver bullet for solving spatial audio problems. No system (Ambisonics, VBAP, etc.) is in itself a solution, especially for the sonic artists who want to employ the full range of spatial audio. And, every situation needs to be understood in appropriate terms. Possibly the most artistically successful implementations of spatial audio are yet to come.

6. REFERENCES

[1] Adair, S., Alcorn M. and Corrigan, C. "A Study Into the Perception of Envelopment in Electroacoustic Music," *Proceedings of the 2008 International Computer Music Conference,* Belfast, UK, ICMA, 2008.

[2] Blauert, J. "Localization and the Law of the First Wavefront in the Median Plane," *Journal of the Acoustical Society of America* 50(2), 1971, pp. 466-70.

[3] Blauert, J. *Spatial Hearing: The Psychophysics of Human Sound Localization*, MIT Press, Cambridge, MA, 1983.

[4] Bloom, P. "Creating Source Elevation by Spectral Manipulation," *Journal of the Audio Engineering Society* 25, 1977, pp. 560-5.

[5] Bradley, J. and Soulodre, G. 1998. "Objective Measures of Listener Envelopment," *Journal of the Acoustical Society of America* 98, 1998, pp. 2590-2597

[6] Gardner, W. *3-D Audio Using Loudspeakers*, Ph.D. Dissertation, MIT, 1997.

[7] Haas, H. "On the influence of a single echo on the intelligibility of speech," *Acustica* 1, 1951, pp. 49-58.

[8] Holman, T. *Surround Sound Up and Running.* 2nd Edition. Amsterdam, Elsevier, 2008.

[9] Kendall, G. "The Decorrelation of Audio Signals and Its Impact on Spatial Imagery," *Computer Music Journal,* 19(4), 1995.

[10] Kendall, G. and Ardila, M. "The Artistic Play of Spatial Organization: Spatial Attributes, Scene Analysis and Auditory Spatial Schemata," *Proceedings of the 2007 International Computer Music Conference.* Copenhagen, ICMA, 2007.

[11] Kendall, G. "Spatial Perception and Cognition in Multichannel Audio for Electroacoustic Music," *Organised Sound*, 15(3), 2010.

[12] "Listen HRTF Database," IRCAM, http://recherché.ircam.fr/equips/salles/listen/index.html.

[13] Litovsky, R., Colburn, H. S., Yost, W. A. and Guzman, S. "The Precedence Effect," *Journal of the Acoustical Society of America* 106(4), 1999, pp. 1633-54.

[14] Roffler, S. and Butler, R. "Localization on tonal stimuli in the vertical plane," *Journal of the Acoustical Society of America* 43, 1968, pp. 1260-66.

[15] Roffler, S. K. and Butler, R. A. "Factors that influence the localization of sound in the vertical plane," *Journal of the Acoustical Society of America* 43, 1968, pp. 1255-59.

[16] Rumsey, F. "Spatial Quality Evaluation for reproduced Sound: Terminology, Meaning, and a Scene-based Paradigm". *Journal of the Audio Engineering Society* 50(9), 2002.

[17] Schroeder, M. and Atal, B. "Computer Simulation of Sound Transmission In Rooms." *IEEE Int. Conv. Record*, 7, 1963, pp. 150-155.

[18] Susnik, R., Sodnik, J. and Tomazic, S. "Coding of Elevation in Acoustic Image of Space," In *Acoustics: Proceedings of the Annual Conference of the Australian Acoustical Society 2005*, pp. 145-50.

[19] Wallach, H., Newman, E. B., Rosenzweig, M. R. "The precedence effect in sound localization," *American Journal of Psychology* 57, 1949, pp. 315-36.

[20] Wakuda, A., Furuya, H., Fujimoto, K., Isogai, K. & Anai, K. "Effects of arrival direction of late sound on listener envelopment," *Acoust. Sci. & Tech.* 24, 2003, pp. 179-185.