Directional Sound Processing In Stereo Reproduction

Gary S. Kendall
Computer Music Studio, School of Music, Northwestern University, Evanston, IL 60201
(708) 491-3178; gary@music.nwu.edu

1. INTRODUCTION

In everyday life, judging the direction of a physical event in three-dimensional space is dependent on continuous sensory interaction with the environment. When such an event creates perturbations in air pressure, energy from that event is transmitted through the air to the listener who perceives an acoustic event. There is typically one direct sound path and many indirect sound paths due to reflections in the environment. The listener's judgement of the direction of the sound event is dominated by the sound that reaches the listener along the shortest, most direct path. The spatial information reaching the ears is dependent on the acoustic interaction of the sound source with the listener's torso, head, pinna, and ear canal. The composite of these properties can be measured and captured as a "head-related transfer function" (HRTF). HRTFs are highly dependent on the direction of the sound source (Figure 1). Low frequency sound wraps around the head and high frequency sound is blocked by the head producing significant frequency-dependent differences at the two ears.

Even though HRTFs are very rich in acoustic detail, perceptual research indicates that the auditory system is selective in the acoustic information that it utilizes in making judgements of sound direction. The majority of this research focuses on binaural cues (information from the two ears combined), although there is also research into monaural cues (information from the individual ears). Classical psychoacoustics has concentrated on the role of interaural cues (differences between the ears), both interaural intensity difference (IID) and interaural time difference (ITD) (see time domain representation in Figure 2). IID and ITD primarily determine the extent of the lateralization of the sound source, that is, its relative position along a left/right axis. It should also be pointed out that these classic experiments have not utilized the frequency-dependent ITD and IID typical of HRTFs (see frequency domain representation in Figure 2) and have not investigated the perception of source directions on the front/back or up/down axes. Evidence does reveal that monaural temporal information is irrelevant to spatial perception and that interaural temporal information is extremely important. While front/back discrimination is possible on the basis of the full acoustic information in HRTFs, it is also clear that head movement plays a dominant role in resolving front/back differences (Wallach, 1940). There is considerable controversy at present concerning the impact of individual differences in HRTFs which vary tremendously among individuals. It appears that some individual's HRTFs improve other individual's localization accuracy, but that large differences in HRTFs of individuals can undermine localization. At the same time it appears that effective localization can occur in many cases in which the ears receive directional transfer functions (DTFs) that bear little resemblance to measured HRTFs. Kendall and Rodgers (1982) used low-order filters to create cartoon-like approximations of natural HRTFs while Martens (1987) and Kendall et al. (1988) describe the use of principal components analysis to create artificial DTFs.

2. STEREO REPRODUCTION OF DIRECTIONAL CUES

In typical stereo reproduction of directional cues, a monophonic source input signal bifurcates to form a stereo pair, each channel is processed by a directionally dependent digital filter, and the processed stereo pair reproduced by the headphones or loudspeakers. Pairs of measured HRTFs or DTFs can be directly implemented as FIR digital filters, and a library of filter coefficients with a dense sampling of directions can be stored in computer memory (Figure 3). (Not considered here are computational techniques for environmental simulation affecting the perception of distance and spaciousness that are discussed by Kendall and Martens, 1984, and Kendall et al., 1986.)

A. HEADPHONES. It would seem intuitively obvious that headphone reproduction provides the most controlled method for the reproduction of directional cues, but the task is far more difficult than one might expect. In order to avoid changes in timbral color, the headphone system must be equalized in order to compensate for the acoustic properties of the transducers and the coupling to the ears. Typically the measured transfer function of the headphone system is divided out prior to reproduction. The response of headphone transducers vary from one model to another and tend to be quite deficient in high and/or low frequencies. These deficiencies cannot be completely compensated by equalization. The coupling to the ears changes with each reseating of the headphones and therefore no one measurement is sufficient (Figure 4). It is recommended that the equalization function be calculated through critical-band spaced smoothing of the measured spectra and averaging of representative measurements before inverting the response. Headphone reproduction of traditional stereo recordings typically creates the impression that sound events are originating inside the head with a bias toward the rear. These difficulties must be overcome in reproducing directional cues: sound images must be perceived outside the head ("externalization") and frontal images must not be confused with back images ("front/back discrimination"). Externalization is aided by the presentation of ambient sound with interaural incoherence and front/back discrimination can be improved through modifications of HRTFs.

Head Tracking. The problem of creating front/back discrimination is largely solved when headphones are coupled to a head tracking system that senses the listener's head orientation and position. A computer receiving this information can continuously update the directional filters in order to maintain the absolute position of the sound source within the environment as the listener's head moves. This simulates the natural kind of interaction the listener has with the environment. Because the auditory system is

very sensitive to time lag in the change of the directional cues due to the head turns, it is very important to minimize the latency of response of the headtracker and computer.

B. LOUDSPEAKERS. In stereo reproduction with loudspeakers, the seating position of the listener and the distance between the loudspeakers have tremendous impact on the potential spatial imagery. When the two stereo loudspeakers produce the same acoustic signal (regardless of whether HRTFs are present), only listeners positioned equidistant from the loudspeakers hear the sound image positioned between the loudspeakers. The apparent location of the sound image shifts toward the leading loudspeaker if there is delay to one ear due to the listener sitting off center and is firmly located at the leading loudspeaker with delays around 1 msec. This is due to the "precedence effect" (Wallach, et al., 1949) by which the auditory system gives preference to the first arriving sound (the auditory system views the signal from the second loudspeaker just as it would a room reflection).

Large-space reproduction. Increasing the distance between the loudspeakers as occurs in concert halls or theaters increases the range of seating locations over which the absolute time of arrival difference causes the apparent sound image to collapse into the closer loudspeaker. Auditory localization is relatively robust in the range of time differences associated with the size of the head, but completely overwhelmed at time delays commonly experienced with loudspeakers in large rooms. This virtually rules out the use of directional transfer functions and suggests multiple loudspeakers when directionalizing sound images in large rooms.

Near-field reproduction. Illusions of sound direction will be most successful when the listener's position relative to the loudspeakers is fixed and known in advance as can occur in near-field reproduction settings such as living rooms and audio control rooms. Figure 5 shows an idealized loudspeaker reproduction setting and illustrates the paths by which sound reaches the listener's ears. Loudspeaker reproduction is similar to headphone reproduction in that sound emanating from the left loudspeaker arrives at the left ear and sound from the right loudspeaker arrives at the right ear and that both signals must be equalized. The two acoustic signals arriving at the ears have superimposed on them the HRTFs for the loudspeaker direction relative to the ipsilateral ear (typically 30-degrees off from straight ahead in the horizontal plane, H_{30}). If the source material already includes HRTFs, it must be equalized so as to remove the H_{30} cue. Environmental reflections of sound arriving within 1 msec will corrupt the HRTFs. Therefore sound reflections near the loudspeakers or listener must be eliminated.

Cross-talk. There are also acoustic signals that reach the ears from the loudspeakers on the other side of the head, for example, the signal from the left loudspeaker arrives at the right ear (Figure 5). These signals have superimposed on them the HRTF for the loudspeaker direction relative to the contralateral ear (typically 330-degrees in the horizontal plane, H_{330}). These signals reaching the ears on the opposite side from each loudspeaker are typically referred to as acoustic "crosstalk." Crosstalk is present in all stereo reproduction. Even in the best of reproduction settings, crosstalk has an impact of sound coloration. Figure 6 shows the change in magnitude response at the ears that results from cross-talk and the deep notch created around 2K Hz. Even though we are accustomed to the presence of cross-talk and typically ignore it, one can learn to hear it in a reproduction environment that is free of room reflections.

Cross-talk Cancellation. The first significant loudspeaker reproduction system for directional was achieved by Schroeder and Atal (1963) and, despite the early date, it has served as the foundation for most loudspeaker systems ever since. In order to deliver to the ears the HRTFs associated with an illusory source location, this system has both to equalize the HRTF for the loudspeaker location and to eliminate the cross-talk signals. It eliminates the cross-talk signals by issuing from the near loudspeaker a signal that could acoustically cancel the cross-talk signal from the far loudspeaker. This is represented in Figure 5. (The system is actually a bit more complex than described here.) The Schroeder-Atal system has many descendants the best of which is the system described by Cooper and Bauck (1988).

All of the variants of this system are constrained by a set of assumptions that produce practical limitations. Just as with headphones, because there are individual differences in HRTFs, equalization is seldom perfect. This becomes particularly problematic with the cancellation signal which must match the 330° HRTF. Most importantly in order to cancel the high frequency content of the HRTFs, there must be an exact match between the signals arriving at the head and the cancellation signal. This is undermined by individual differences in HRTFs. In fact, crosstalk cancellation systems seldom cancel high frequency information which is typically localized toward the loudspeakers even when the low-to-mid-range content is localized toward the side or rear. Small variances in the head position relative to the loudspeakers can cause total phase reversals of the cancellation signal and dense combing. It is typical that a shift in head position of less than nine inches will totally collapse the imagery. All of the known crosstalk cancelation systems also explicitly assume that the auditory system requires natural HRTFs at both ears. This appears to be an unnecessary assumption (tantamount to accepting that directional hearing cues are monaural).

Alternative Approaches. An alternative to this approach was reported by Kendall and Rodgers (1982), who describe achieving significant loudspeaker location with low-order digital filters that provided simple approximations of HRTFs without the benefit of crosstalk cancellation. (In retrospect, it appears that the salience of this system was due to the interaural phase relationships, not to the HRTF approximations.) Another alternative was achieved by Lowe and Lees (1991) who took a purely empirical approach and constructed very effective directional transfer functions by direct experimentation with gated sinusoids (thereby capturing interaural group delay). Some of the same problems associated with crosstalk cancellation affect these alternative approaches as well. Variances in head position cause inaccuracies in the high frequency information arriving at the ears (because crosstalk is never eliminated, the left and right loudspeaker signals combine acoustically at the ears and cause phase

shifts and cancellations). The primary advantages are that these systems are less sensitive to the listener's seating location. Kendall and Martens report that circular sound paths retain their general shape and deform in a graceful manner even as the listener moves far off center. Lowe and Lees report that listeners were able to rotate their heads and orient toward the sound sources. Even with these alternative approaches, the loudspeaker reproduction environments often inhibit the creation of images in one or more spatial regions due to early reflected sound in the reproduction environment and/or asymmetries in the reproduction equipment. Most susceptible are rear images which often shift to the front or cling close to the listener's head and side images that collapse toward the front due to shifts in the location of the listener's head.

3. CONCLUSION

Both headphone and loudspeaker reproduction of directional cues present tractable problems and can be very successful in controlled reproduction settings. Headphone reproduction with head-tracking provides the most resilient form of reproduction but it is also the most complicated and expensive due to the overhead of dynamic filtering and head-tracking. Loudspeaker reproduction, even when limited to near-field monitoring, is more convenient but less resilient than headphones. As the technology for reproducing directional cues becomes increasingly refined (and less expensive), different technical issues begin rise to the surface. Increasing the level of complexity from reproducing single directional cues to reproducing full spatial environments necessitates a tremendous increase in computational bandwidth. Simulated natural environments must be able to contain many individual sound sources and to replicate the reflected sound arriving at the listener from all directions. Many pairs of dynamic directional filters are required. Then too, in everyday life, people are *interactively* involved with their environment and themselves initiate many of the sound events. Without interaction, listeners are only empowered to sit passively as sound sources move by or to fly as a disembodied spirit around events that can be heard but not touched. The computational requirements of interactive spatial sound are tremendous but must be met if we are to simulate a full virtual acoustic reality.

4. ACKNOWLEDGMENTS AND REFERENCES

Appreciation and thanks to Bill Martens and Marty Wilde for many years of support and friendship while studying auditory localization. Thanks to Ric Ashley and Amnon Wolman for their personal support in helping me to get restarted at Northwestern.

Cooper, Duane H. and Jerald L. Bauck (1988). "Prospects for Transaural Recording" Preprint #2734, 85th Convention of the Audio Engineering Society. November, 1988.

Kendall, G. S. and C. A. P. Rodgers (1982) "The simulation of three-dimensional headphones cues for headphone listening" *Proceedings of the 1982 International Computer Music Conference.*

Kendall, G. S. and W. L. Martens (1984) "Simulating the Cues of Spatial Hearing in Natural Environments" *Proceedings of the 1984 International Computer Music Conference.*

Kendall, G., W. Martens, D. Freed, D. Ludwig, and R. Karstens (1986). "Image Model Reverberation from Recirculating Delays" Preprint # 2408, 81st Convention of the Audio Engineering Society. December, 1986.

Kendall, Gary S., William L. Martens, and Martin D. Wilde (1988). "A Spatial Sound Processor For Loudspeaker and Headphone Reproduction" *The Sound of Audio. Proceedings of the AES 8th International Conference.*

Lowe, Danny D. and John W. Lees (1991). "Sound Imaging Process" U.S. Patent #5,046,097.

Martens, William L. (1991). *Directional Hearing on the Frontal Plane: Necessary and Sufficient Spectral Cues.* PhD dissertation, Northwestern University.

Martens, William (1987). "Principal components analysis and resynthesis of spectral cues to perceived direction. *Proceedings of the 1987 International Computer Music Conference.*

Schroeder, M. R., and B. S. Atal. (1963). "Computer Simulation of Sound Transmission in Rooms" *IEEE Conv. Record*, pt. 7, pp. 150-5.

Wallach, Hans, Edwin B. Newman, and Mark R. Rosenzweig (1949). "The Precedence Effect in Sound Localization" *Journal of Experimental Psychology* 32 (3), pp. 315-36.

Wallach, Hans (1940). "The Role of Head Movements and Vestibular and Visual Cues in Sound Localization" *Journal of Experimental Psychology* 27 (4), pp. 339-68.

5. FIGURES

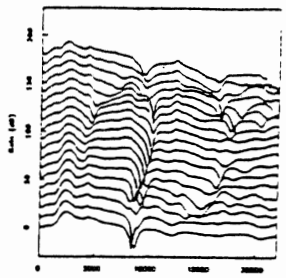

Figure 1. Magnitude spectra of HRTFs measured at the ipsilateral ear for directions on the horizontal plane from 0° (directly in front) to 180° (directly behind).

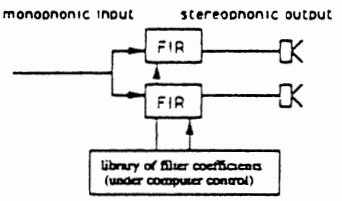

Figure 3. FIR digital filter implementation of HRTFs under computer control.

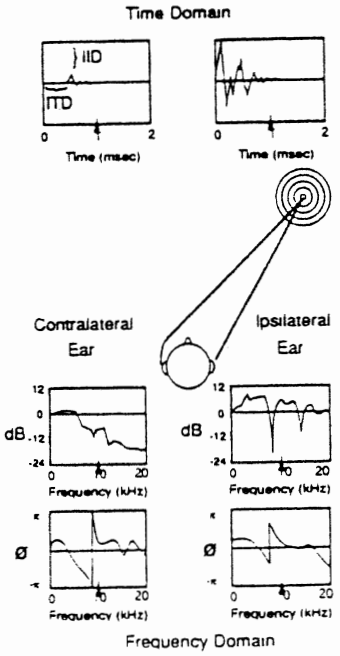

Figure 2. Time domain and frequency domain representations of HRTFs for the ipsilateral and contralateral ears.

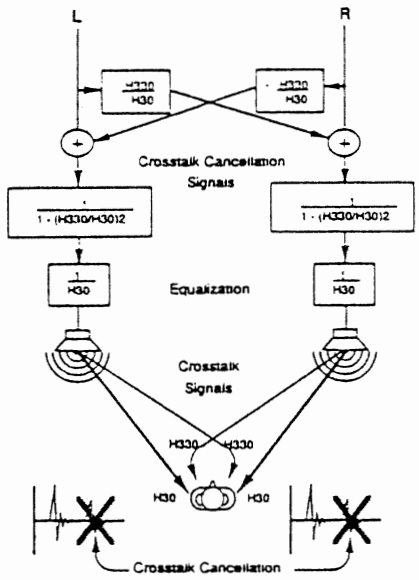

Figure 5. Acoustic crosstalk in loudspeaker reproduction and the Schroeder-Atal method for crosstalk cancellation.

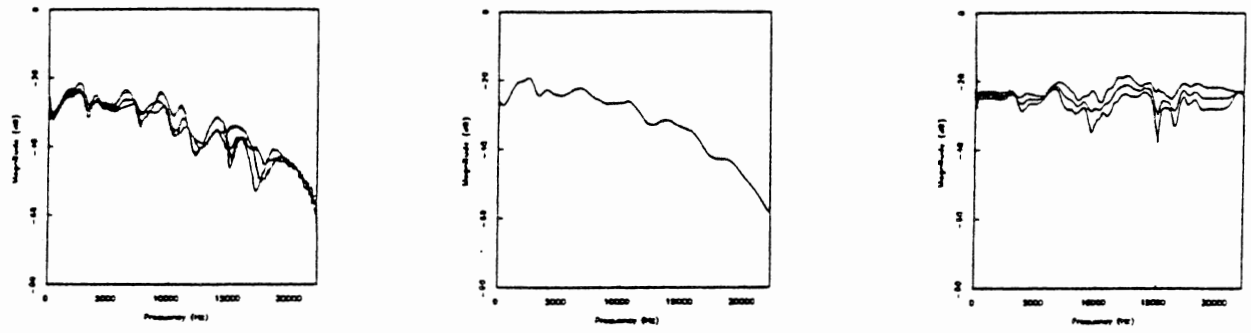

Figure 4. Headphone equalization: (a) magnitude response measured for five reseatings of STAX SR Lambda earphones; (b) critical-band smoothed, mean magnitude function which is inverted for equalization; (c) magnitude response of reseatings measured with equalization (mean and one standard deviation above and below). Reproduced from Martens (1991).

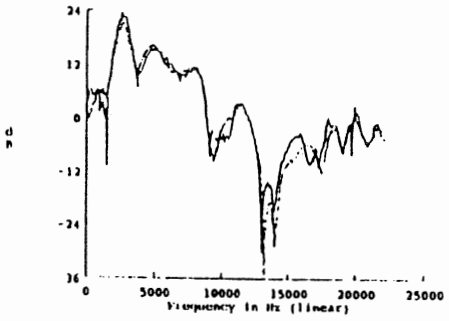

Figure 6. Magnitude response measured at listener's right ear in stereo loudspeaker reproduction. Dotted line: one loudspeaker on ipsilateral side; solid line: two loudspeakers with crosstalk.