
Gary S. Kendall

Center for Music Technology
School of Music
Northwestern University
Evanston, Illinois 60208, USA
g-kendall@nwu.edu

The Decorrelation of Audio Signals and Its Impact on Spatial Imagery

Background

As used here, the term “decorrelation” refers to a process whereby an audio source signal is transformed into multiple output signals with waveforms that appear different from each other, but which sound the same as the source. In the experience of most sound professionals, decorrelation occurs as a by-product of other acoustic or electronic processes that often change the sound of the source. In acoustic performances, decorrelation occurs as a by-product of reverberation and chorusing, and in digital signal processing, “stereoized” reverberation and chorusing achieve the same effect. Decorrelation occurs in sound synthesis when there are slight differences between the sounds synthesized for the output channels. That often happens with granular synthesis, but can also happen with frequency modulation or additive synthesis if the composer takes special care in designing the algorithms. In the audio industry, there is a long tradition of devices for the home or studio that “stereoize” monophonic signals, and they too typically decorrelate the output channels. Numerous settings on effects processors for flanging, combing, etc. produce decorrelated output. In recording studios, vocal artists sometimes are recorded twice on separate tracks so that the micro-variations in the two performances create decorrelation.

Why focus on decorrelation as a separate aspect of these processes? In the field of spatial hearing, signal decorrelation is known to have dramatic impact on the perception of sound imagery. The degree to which sounds are decorrelated has proven to be a significant predictor of perceptual effects, both in natural environments and in audio reproduction.

Therefore, all of the diverse processes mentioned above are related to each other by the impact of decorrelation on the spatial imagery of the sound. While there is a considerable literature on spatial sound processing, this literature is usually concerned with one of two goals: (1) positioning sound images at a particular location in three-dimensional space, or (2) creating three-dimensional simulated environments. These goals are important, but there are obviously many other creative potentials for spatial sound processing, and other kinds of practical problems to solve. For example, decorrelation can produce sound images with the width, depth, and spaciousness typical of natural environments while circumventing the computational burden of a full environmental simulation.

In audio reproduction, decorrelation has at least five effects on the perception of spatial imagery:

1. The timbral coloration and combing associated with *constructive and destructive interference* of multiple delayed signals is perceptually eliminated.
2. Decorrelated channels of sound produce *diffuse sound fields* (akin to the late field of reverberant concert halls).
3. Decorrelated channels produce *externalization* in headphone reproduction.
4. The position of the sound field does not undergo *image shift* with changes in the position of the listener relative to stereo loudspeakers.
5. The *precedence effect*, which causes the collapse of the image into the nearest loudspeaker, is defeated, enabling one to present the same sound signal from multiple loudspeakers.

The discussion that follows is organized in two sections. The first section discusses signal pro-

cessing techniques that create decorrelated signals through direct means. The second section elaborates on the five categories of perceptual effects mentioned above. Also included is an appendix on a related signal processing technique for controlling the perceived distance of a sound image in near-field loudspeaker reproduction.

Much of the work described in this paper was performed during the period from 1988 through 1990 by Marty Wilde, William Martens, and myself in close collaboration both at Northwestern University and at the now-defunct Auris Corporation. More recent work has been completed at Northwestern in collaboration with Matt Moller. Our exploration into the effects of decorrelation on stereo imagery sometimes culminated in simple, concise audio demonstrations; at other times, the complexity of the relationship between acoustics and perception compelled us to run perceptual experiments. Some issues were left unresolved and some parts of the work were left unfinished. The purpose of this article is to survey and summarize work that has not been reported in print, so that others may begin to use these techniques and make further contributions to our community's knowledge.

Techniques for Directly Creating Decorrelated Signals

Some of the ways that decorrelation is produced as a by-product have already been mentioned, but these are not the only ways of creating decorrelated signals. Additional digital signal processing techniques are described below that produce decorrelation directly. With these techniques one can create multi-channel replicas of a given source signal that sound the same as the source and that are nonetheless decorrelated—that is, there are no audible effects other than the spatial effects due to the decorrelation. These techniques enable one to set the level of correlation between any two audio channels in a continuous range from -1.0 to $+1.0$, or to produce an unlimited number of output channels with nearly zero correlation.

Definition of Correlation Measure

The correlation measure of two signals, $y_1(t)$ and $y_2(t)$, can be determined by computing the cross-correlation function, $\Omega(\Delta t)$:

$$\Omega(\Delta t) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^{+T} y_1(t) y_2(t + \Delta t) dt,$$

where Δt represents a temporal offset between $y_1(t)$ and $y_2(t)$. For the purposes of most discussions, the correlation measure (also called the cross-correlation coefficient) is expressed as a single number, and is taken to be the value of the peak in the cross-correlation function with the greatest absolute value. As illustrated in Figure 1, if $y_1(t)$ and $y_2(t)$ are identical, there will be some value of Δt at which they will have the highest possible positive correlation measure, $+1.0$. If $y_1(t)$ and $y_2(t)$ are identical except for being 180 degrees out of phase, there will be some Δt at which they will have the highest possible negative correlation measure, -1.0 . If $y_1(t)$ and $y_2(t)$ are very dissimilar, they are said to be "uncorrelated" and their correlation measure will be near 0 for all Δt .

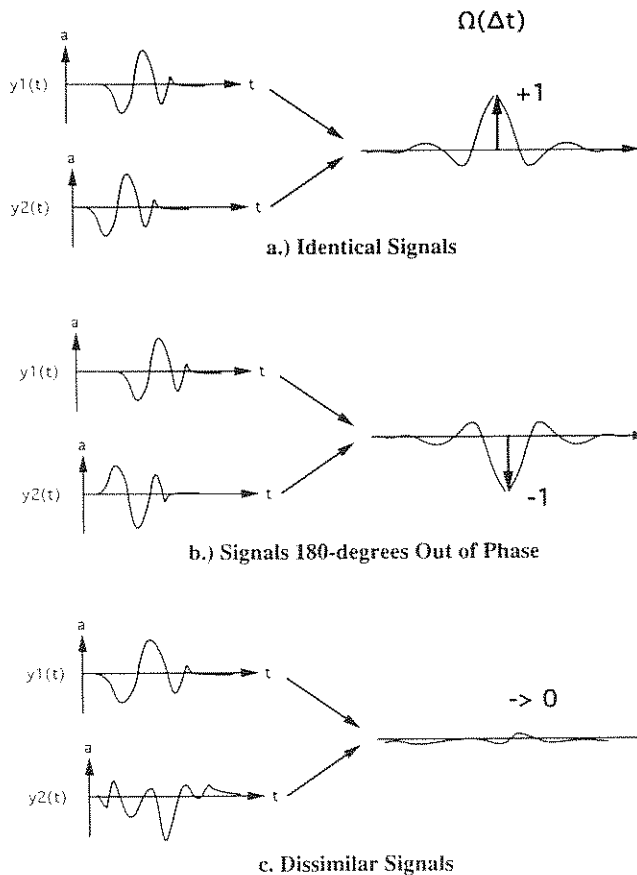
Building Decorrelation Filters

Converting Input to Decorrelated Output

In most of the practical applications of decorrelation, the input will be a monophonic signal (or a multi-channel signal summed to form a monophonic input signal). The user will specify the correlation measure for each pair of output signals in a range from $+1.0$ through -1.0 . For many applications, the optimal correlation measure is 0, and for some of these applications there can be multiple decorrelated output channels.

The easiest way to conceptualize the creation of decorrelated signals is through convolution. To produce a pair of output signals with a specified correlation measure, an input signal can be convolved with each of two exemplar signals that are correlated with each other by the specified amount. This

Figure 1. Cross-correlation function, $\Omega(\Delta t)$, and correlation measure for (a) identical signals, (b) signals 180 degrees out of phase, and (c) dissimilar signals.

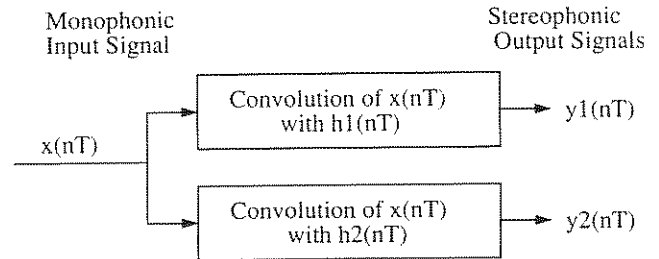


convolution operation itself can be re-envisioned as a finite-impulse-response (FIR) filter and the exemplar signals as the filter's coefficients. The signals resulting from the convolution will be correlated at a level close to that of the exemplar signals.

An illustration of this technique for use with a monophonic input and stereo output is shown in Figure 2. The digital input signal, $x[nT]$, is applied to a pair of FIR filters with coefficient sequences, $h_1[nT]$ and $h_2[nT]$. The output of the FIR filters, $y_1[nT]$ and $y_2[nT]$, is the convolution (denoted by " $*$ ") of the input signal with each coefficient sequence:

$$y_1[nT] = x[nT] * h_1[nT] \text{ and} \\ y_2[nT] = x[nT] * h_2[nT].$$

Figure 2. Decorrelation through convolution. Monophonic input and stereo output.

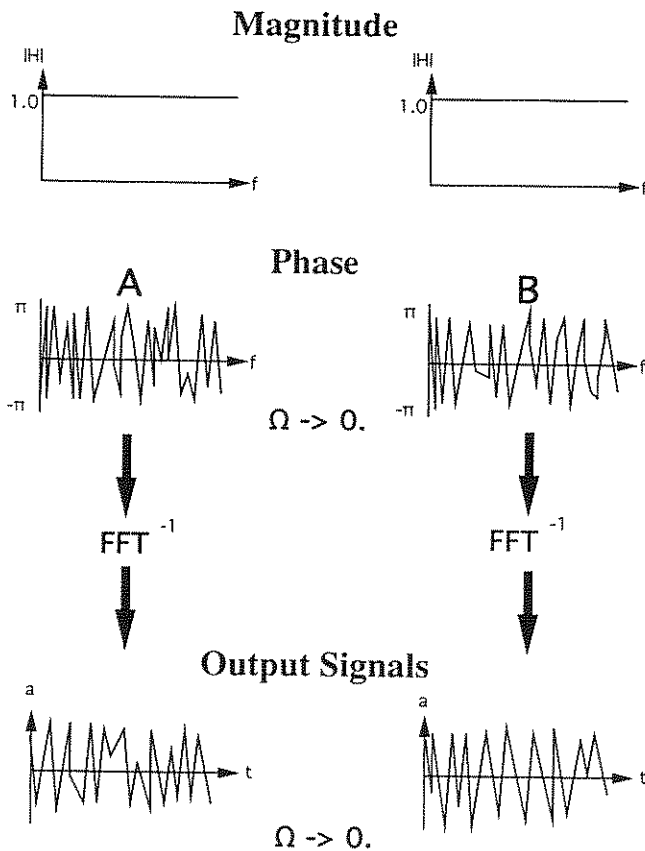


Building a Library of Paired FIR Filter Coefficients

The correlation measure of the output signals is determined by the correlation measure of the FIR filter coefficients. To provide a complete range of correlation measures, a library of coefficients for the paired filters must be created. The filter coefficients are computed from a frequency-domain specification of both magnitude and phase via the inverse Fast Fourier Transform (IFFT), as shown in Figure 3. The magnitude part of the specification will be set to unity across all frequencies, that is, all of the FIR filters will be all-pass. The phase part of the specification will be constructed from combinations of random number sequences. The resulting correlation measures are determined only by the phase information and, in fact, the entire range of possible correlation measures is attained merely through inter-channel phase manipulations! If the correlation measure of the phase sequences is near 0, the correlation measure of the time sequences produced by the IFFT will be near 0 also. Output signals will preserve the timbre of the input source because the filters are all-pass, and also because the phase of a single-channel signal has no impact on the perception of timbre with the exception of a few special circumstances (Plomp and Steeneken 1969). While single-channel phase has little impact on timbre, inter-channel phase has a dramatic impact on spatial perception.

To construct a complete library of paired filter coefficients, start with two independent random number sequences (A and B) whose amplitude values are scaled to the range of $+\pi$ to $-\pi$. The process of creating the library varies with the specified correlation measure, Ω' . The library can be constructed in

Figure 3. The computation of coefficients for a pair of FIR filters from a frequency domain specification via the IFFT. The magnitude is set to unity, and the phase is constructed from random number sequences A and B. The correlation measure of the coefficient sequences will approach zero.



the following steps (paraphrased from Wilde 1989 and Wilde et al. 1989):

1. $\Omega' = +1$. Only the A-scaled sequence is used as the phase specification for both the left and right channels. Coupled with each of these identical left- and right-channel phase specifications is a unity magnitude specification. The IFFT is then applied to each of these two pairs of magnitude and phase specifications in the frequency domain, to generate two sets of FIR coefficients, $h_1(nT)$ and $h_2(nT)$. The correlation measure of $h_1(nT)$ and $h_2(nT)$ is $+1.0$, and the two output signals will be identical.
2. $\Omega' = -1$. Again, only the A sequence is used as the phase specification for the left channel, while π is added to that same sequence and used as the right-channel phase specification.

These two phase specifications are each coupled with a unity magnitude, and the IFFT is applied to each of these two pairs of magnitude and phase specifications to create the two sets of FIR coefficients, $h_1(nT)$ and $h_2(nT)$. The correlation measure of $h_1(nT)$ and $h_2(nT)$ is -1.0 , and the two output signals will be 180 degrees out of phase.

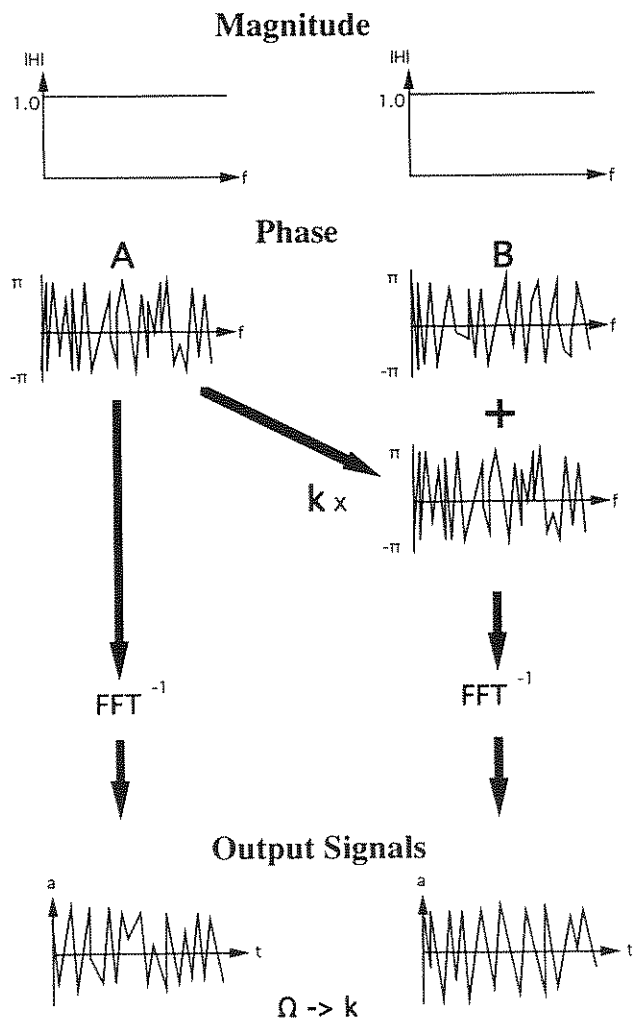
3. $\Omega' = 0$. The scaled A sequence is the left-channel phase specification, and the scaled B sequence is the right-channel specification. Coupling each of those sequences with a unity magnitude spectrum and applying the IFFT to both pairs of frequency domain specifications yields the two sets of coefficients. The correlation measure of $h_1(nT)$ and $h_2(nT)$ is very near to 0. (See Figure 3.)

For all remaining correlation levels, the two scaled sequences A and B are first summed to form the left-channel phase specification. (Note that if a phase value should exceed $|\pi|$, it is "wrapped" back around into the range from $-\pi$ to $+\pi$.)

4. $0 < \Omega' < 1$. Sequence A is weighted by a scaling coefficient, k , before being summed with the full-scale B sequence to become the right-phase specification. The value of k is limited to the range $0 < k < 1$, and is dependent upon the desired output correlation level. Coupling these two phase specifications with unity magnitude spectra, and applying the IFFT to each of the pairs of frequency domain sequences yields two sets of coefficients. The correlation measure of $h_1(nT)$ and $h_2(nT)$ is very near to $+k$. This is illustrated in Figure 4.
5. $0 > \Omega' > -1$. Sequence A is again weighted by some scaling coefficient, k . But unlike the procedure for positive correlation levels, π is added to the A sequence before being summed with the B sequence to become the right phase specification. Coupling these two phase specifications with unity magnitude spectra and applying the IFFT to both pairs of the frequency domain specifications yields two sets of coefficients. The correlation measure of $h_1(nT)$ and $h_2(nT)$ is very near to $-k$.

Figure 4. The computation of coefficients for a pair of FIR filters by summing the k -weighted A sequence with the B sequence in

one channel. The correlation measure of the resulting coefficient sequences will approach k .

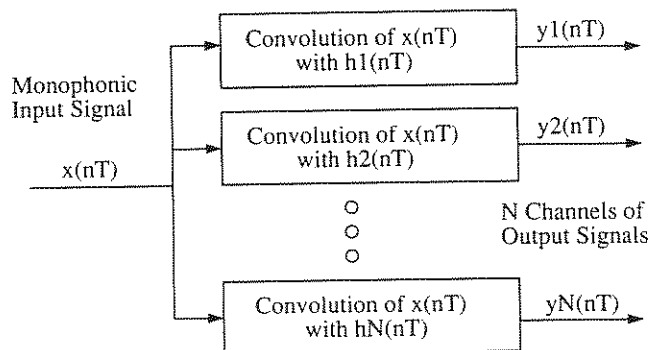


There are alternative methods for creating number sequences with constant magnitude and random phase. See Manfred Schroeder (1984) for one such approach.

Building a Library of Multi-Channel FIR Filter Coefficients with Correlation Measures Near Zero

It is easy to see from the above description that three or more sets of filter coefficients with correlation measures near 0 can be constructed as in step 3, by using three or more independent random number sequences. A library designed to support N channels of output will contain coefficients for N

Figure 5. Decorrelation with monophonic input and multiple output channels.



filters. The library is created by starting with N independent random number sequences $\{A_1, A_2, \dots, A_N\}$ in place of A and B . An illustration of this technique for use with a monophonic input and multiple output channels is shown in Figure 5. The correlation measure of the output of any pair of filters will be very near to 0.

Practical Limitations and Perceptual Concerns

The filter design method discussed so far attempts to avoid any alterations in the timbre of input sound by maintaining constant magnitude across frequency. This is not as easy as it first appears. The points specified in the frequency domain for magnitude and phase are linearly spaced in frequency, and the magnitude spectrum that results from using the IFFT to produce the FIR coefficients will not be constant in between the specified frequency points, as shown in Figure 6a. Therefore, one expects that timbral neutrality is improved by specifying a higher number of points and producing a higher number of coefficients. However, the number of coefficients is approximately twice the number of points specified in the frequency domain, and the temporal duration of the filter's impulse response must be shorter than around 20 msec to avoid diffusion in the time domain, which would smear the transient properties of the input signal.

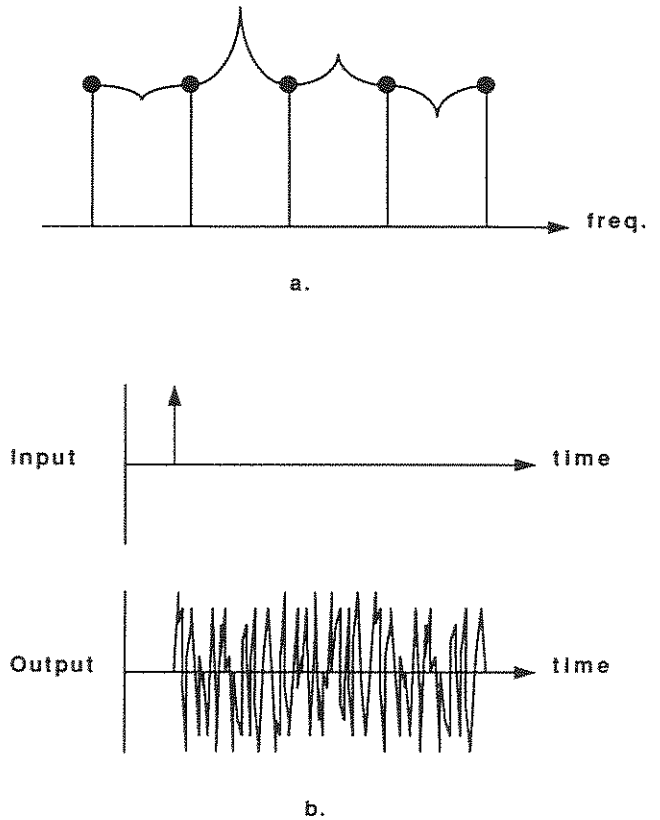
As shown in Figure 6b, the energy of an impulse becomes evenly spread over a duration determined by the number of coefficients. Timbre obviously depends not only on the spectrum, but also on

the temporal evolution of the source signal. Consider, too, that the magnitude of the potential phase shift on low-frequency components of the input signal is diminished by decreasing the number of coefficients. Consequently, for any given sampling rate, there is a tradeoff between timbral neutrality and the impact at low frequencies. Our practical experience has shown that sound sources containing the most transient information (such as speech) should be processed with fewer coefficients than most other sound sources (such as music). Experience has also shown that timbral coloration is less noticeable when decorrelation is applied to the individual tracks rather than to an entire mix.

Another limitation on the filter design is that the finite length of the random number sequences causes imprecision in the match of the prescribed correlation measure to that measured with the output signals. A practical solution was found by generating several candidate filter pairs with different root random-number sequences, and selecting the pairs that produced the best match to the prescribed correlation measures. Then, too, when the input is processed so as to create a correlation measure near 0 or within the range between -0.4 and $+0.4$, the actual cross-correlation function may exhibit positive and negative peaks with similar absolute magnitude. The auditory system does not discriminate very well among correlation measures near 0 (Pollack and Tritpoe 1959), and so the variance between prescribed and measured correlation is of little consequence. The auditory system easily discriminates among correlation measures near $+1.0$ and -1.0 , and here the match between prescribed and measured correlation is quite good.

Another consideration with impact on perception is the spacing of the filter's phase characteristics across frequency. A straightforward implementation of the design technique with FFTs leads to spacing that is linear with respect to frequency. An alternative is to space the filter's phase characteristics in log frequency or to model critical band spacing. This is accomplished by associating the random phase values with selective frequencies and

Figure 6. Limitations of the filter design technique: the magnitude spectrum will not be constant in between the specified frequency points (a), and FIR filters cause the energy of an impulse to become evenly spread over a duration determined by the number of coefficients (b).



then interpolating the points in between. Experience once again has suggested that some effects are improved by critical band spacing (for example, diffuse sound fields, especially the high-frequency range), while others work best with the denser, linear spacing (eliminating constructive and destructive interference).

Once the method is chosen for generating a particular number of coefficients at a particular correlation level, it is still possible to generate many unique sets of candidate coefficients, each with a different random number seed. Most sets of coefficients will sound the same or have very subtle differences, but occasionally one will be less effective. This is to be expected, given the finite length of the random number sequences used in creating the coefficient sets. In the end, each candidate set of filter coefficients should be evaluated by ear.

Figure 7. Pole-zero design technique for all-pass filters with randomization of pole distance from the unit circle (r_1, r_2, r_3, \dots are random numbers between 0 and 1).

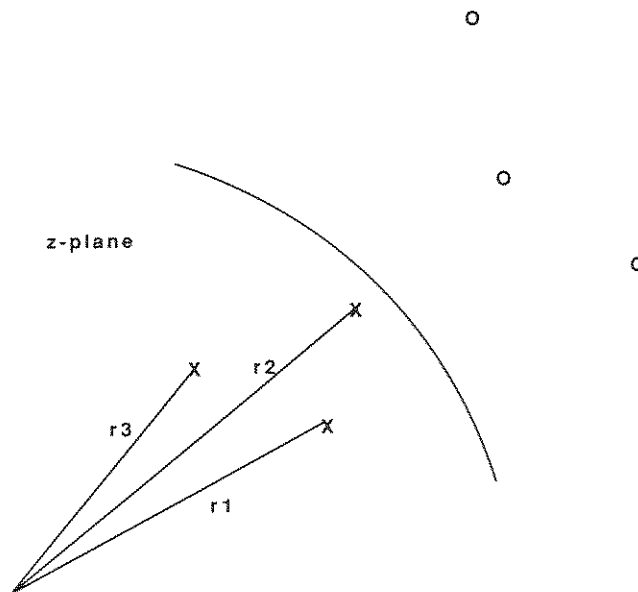
Alternative Approaches

Infinite-Impulse-Response (IIR) Filter Design

An alternative approach to designing all-pass filters is to use pole-zero techniques that are well described in the literature. One such approach is illustrated in Figure 7, wherein the distance of poles and zeros from the unit circle in the z-plane is controlled by a random sequence. Experience has shown that the order of these IIR filters must be high (approaching several hundred) to produce a correlation measure close to 0. At such high orders, finite word-length effects begin to produce discrepancies from an all-pass response that resemble those seen with the FIR approach discussed above. Nevertheless, the use of IIR filters appears to reduce the potential for timbral coloration. This may be partly due to the phase response of these filters, which does not exhibit a flat distribution of phase between $-\pi$ and π (and which causes the impulse response to concentrate energy at one point in time).

Dynamic Decorrelation

There is a significant advantage to IIR decorrelation filters over FIR filters, in that they more easily permit dynamic variations. The IIR filter's coefficients can be continuously updated by randomly varying the distances of the poles and zeros from the unit circle. A new set of coefficients can be easily computed from the new pole/zero locations. In the case of FIR filters, continuous updating of the coefficients is possible, but only at the expense of adding a huge computation burden in calculating the IFFT for each new set of coefficients! Dynamic IIR filters are more practical than FIR filters for real-time applications. We also observe that dynamic variation produces a spatial effect akin to the sound of an environment with moving reflecting surfaces or moving sound sources, such as the movement of leaves and branches in a forest or the movement of a crowd within an auditorium. Dynamic decorrelation imparts a quality of liveliness to a sound field that is missing in the FIR implementation.



The Effects of Multi-Channel Signal Decorrelation in Audio Reproduction

The introduction to this article listed five effects of decorrelation on the perception of spatial imagery. They will now be discussed in depth.

Effect No. 1: Elimination of the Perception of Constructive and Destructive Interference

Constructive and destructive interference may affect listening in a variety of audio circumstances. In room acoustics, strong reflections often lead to interference patterns that are perceived as part of the acoustic character of a room. In sound reinforcement, multiple loudspeakers and loudspeaker stacks create interference patterns that can be heard especially clearly when the listener is moving in relationship to the loudspeakers. In both of these cases, acoustic waves from a single sound source (whether acoustic or recorded) arrive at different times and with varying intensities. The composite magnitude spectrum will exhibit spectral peaks and notches that result from the constructive

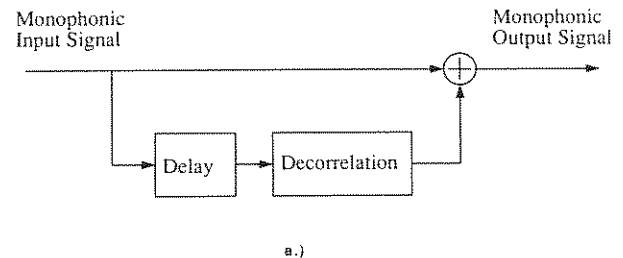
and destructive interference of the acoustic waves. The frequencies of these peaks and notches are dependent on the difference in arrival times of the acoustic signals at the measurement position.

When the arrival of these multiple acoustic waves is integrated into a single perceptual event by the listener, the acoustic constructive and destructive interference gives rise to two interrelated perceptual qualities, "coloration" and "combing." Although these terms are often used by professionals to describe both effects, the meaning of "coloration" will be limited here to changes in the perceived color (or spectral shape) of a sound, and "combing" will be limited to the induction of a pitch whose frequency is the reciprocal of the delay. (This use of the term "combing" is derived from the way recording engineers use it to describe this particular heard quality. The term was originally used to describe the characteristic amplitude response of a "comb" filter.) Coloration can usually be compensated for by changes in equalization, but if the spectral variations are dense, the shape of the spectral envelope may be close to the original, and little change in coloration is perceived. On the other hand, combing seems acute. The auditory system is particularly proficient at picking up the temporal periodicity between the original and delayed signals, from which it creates a pitch percept. Combing is easily detected, and typically difficult to eliminate.

Audio demonstrations have shown that coloration and combing can be eliminated when a delayed signal is decorrelated from the leading signal with a correlation measure approaching 0. The degree to which coloration and combing are removed depends on the correlation measure. Figure 8a shows a signal flow diagram used for creating test signals in which the leading signal is combined with a decorrelated replicant at varying levels of correlation. Figure 8b summarizes the reports of listeners to the test signals. When listeners are presented with a series of test sounds that move from little to complete decorrelation between the leading and replicant signals, the perceived quality of the sound moves from "colored" and "combed" to a restoration of the original's timbre. The shift occurs quickly as the correlation measure approaches

Figure 8. The elimination of coloration and combing in constructive and destructive interference: simplified signal flow diagram for combining original sig-

nal with decorrelated delayed signal (a) and summary of perceptual results depending on the correlation measure (b).



Stimulus	Correlation Measure	Perceptual Result
original alone	∞	"original timbre"
original plus delayed replica: no decorrelation	1.	"colored", "combed"
original plus delayed & decorrelated replica: little decorrelation	.9	"colored", "combed"
complete decorrelation	0.	"original timbre restored"

0. (An unanswered question is how this effect changes over a complete range of time delays.)

An explanation of the phenomenon can be offered by considering that when the decorrelated signal has random phase changes that are spaced more closely than critical bands, the resulting, composite magnitude spectrum will exhibit spectral peaks and notches that are narrower than a critical band, and the critical-band smoothed spectral envelope is likely to retain its original shape. Combing is impossible with a completely decorrelated signal, because it is smeared in time and the temporal periodicity between the original and delayed signal varies with frequency. It is interesting to note that while constructive and destructive interference is still physically present, its perceptual effects are eliminated.

The obvious practical application for the elimination of interference is in loudspeaker reproduction. The implementation should follow Figure 2 for stereo loudspeaker reproduction and Figure 5 for

multiple loudspeaker reproduction. An almost identical implementation was discussed by Augspurger and co-workers (1989), and implemented at the Hollywood Bowl.

Effect No. 2: Creation of Diffuse Sound Fields without Reverberation

Diffuse reverberant sound fields are one of the most important features of concert hall acoustics. The perceived quality of spatial diffuseness is strongly correlated to the “interaural cross-correlation” (or IACC), a statistical measure of the similarity of the acoustic signals arriving at the left and right ears of a listener in the concert hall. A low IACC is strongly correlated to the desired sound quality of “spaciousness” (Schroeder, Gottlob, and Siebrasse 1974; Ando 1977). For the sound reaching the listener directly from the stage, the IACC will be close to +1, meaning that the signals are highly similar (though not identical due to the asymmetry of head acoustics). For the sound reaching the concert-hall listener during reverberation, the IACC will approach 0, meaning that the sound reaching the left and right ears with a separation of just 20 cm is uncorrelated! In fact, almost any point-to-point measurement of reverberation inside the hall would yield similar results—and although the reverberant sound is uncorrelated, it is still clearly from the same source! The impact of the decorrelation is that the sound image does not appear to emanate from any one direction.

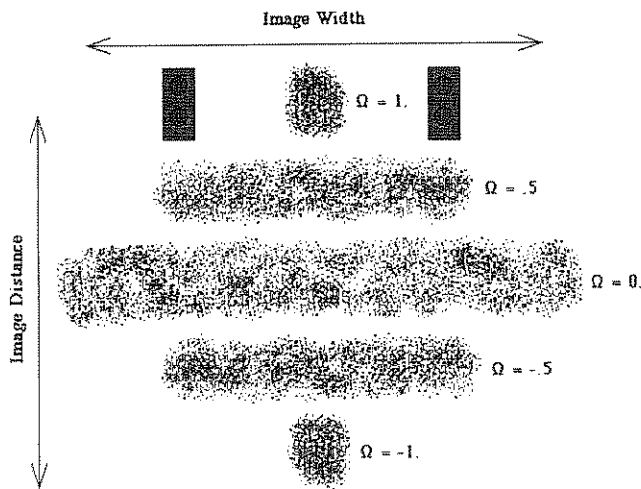
The most commonly used signal processing technique for creating a diffuse sound field is multi-channel reverberation, which mimics the acoustics of a concert hall. Although a spatially diffuse sound field occurs naturally only in the context of reverberation, decorrelation makes it possible through electronic means to create a spatially diffuse sound field without reverberation (acoustic or electronic). For reproduction over loudspeakers, the diffuse sound field is perceived as emanating broadly from around the listener. A complete surrounding image, including the rear, will only occur when the listener is close to the loudspeakers. A common studio technique that achieves a related ef-

fect is to record two versions of the same sound material on separate tracks and reproduce them over separate channels. If the two versions are indeed similar, the listener believes that they represent one performance. However, the micro-differences (essentially, phase differences) between the two versions impede any possibility of forming a single spatial image; the spatial image of the performance is divided between the two channels.

Kurozumi and Ohgushi (1983) supplied an important insight into the impact of IACC on stereo loudspeaker reproduction. They demonstrated that the cross-correlation coefficient of two noise signals presented to listeners over stereo loudspeakers was strongly correlated with two perceptual dimensions: image distance and image width. Image distance is correlated to the value of the cross-correlation coefficient; image width is inversely correlated to the absolute value of the cross-correlation coefficient. For example, the widest image occurs when the cross-correlation coefficient is close to 0; this image is also at a medium distance. The closest sound image occurs when the cross-correlation coefficient is -1.0 , but this also creates a narrow image. The distance and width of sound images is depicted in Figure 9. In addition, Kurozumi and Ohgushi found that the absolute effect of the cross-correlation coefficient is greater for low frequencies (below 1 kHz) than for high frequencies (above 3 kHz).

This was the starting point for a study by Wilde (1989) in which he demonstrated that the image width and image distance of decorrelated natural sound sources was essentially the same as Kurozumi and Ohgushi had found for noise sources. Figure 10 shows Wilde’s multi-dimensional scaling solution from pooled subject data for a string quartet chord, with correlation levels ranging from $+1.0$ to -1.0 in increments of 0.2. Dimension 1 captures image distance, and Dimension 2 captures image width. Thus, to control the width and distance of a diffuse sound field, one can select the filter coefficients associated with the appropriate correlation measure. (See the Appendix, “Controlling Image Distance,” for a further explanation of the relationship among the correlation measure, phase, and image distance.) Because the effect of decorrelation is

Figure 9. Depiction of image width and distance for varying levels of correlation (Ω).



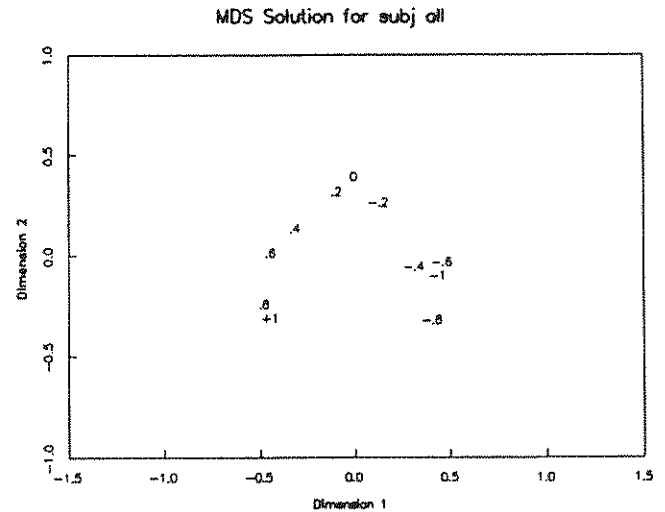
greater for low frequencies than for high, creative applications of decorrelation for diffuse sound fields are most successful when sources have significant low-frequency energy. And although the spatial diffusion does not require reverberation, a single channel of reverberation can be made spatially diffuse by decorrelation.

Effect No. 3: Externalization of Sound in Headphone Reproduction

In everyday life, sound events appear to originate in the environment, but in traditional stereo headphone reproduction, sound events appear to originate inside the listener's head. We have become so accustomed to this effect that it no longer strikes us as bizarre! Externalizing auditory images in headphone reproduction has proved to be an elusive problem. It is especially important in three-dimensional sound, and this is the context in which it has been most thoroughly studied. Externalization is a complex phenomenon that has been found to be affected by a variety of factors including the presence of reverberation (Durlach et al. 1992). As described in the previous discussion of diffuse sound fields, decorrelation is an essential component of reverberation, and appears to be the factor that influences externalization.

Figure 10. Two-dimensional multi-dimensional scaling solution of all pooled subject data for a string quartet chord at correlation levels from

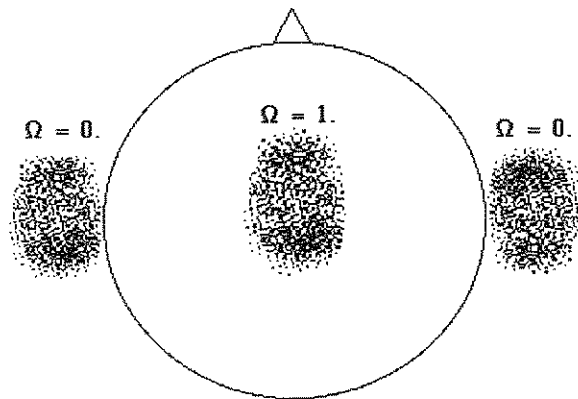
+1.0 to -1.0 in increments of 0.2. Dimension 1 captures image distance, and Dimension 2 captures image width (Wilde 1989).



Some perceptual studies of the effect of correlation level on headphone imagery have asked listeners to report the location of auditory images by drawings (Blauert and Lindemann 1986), and these provide an excellent method for describing how headphone imagery appears to the listener. Figure 11 depicts the informally observed difference in perceived image location for correlated and decorrelated sound sources when decorrelation is achieved with FIR decorrelation filters. The location of decorrelated sources is outside the head to the left and right.

Decorrelation appears to affect the externalization of correlated sources as well. If decorrelated reverberation is added to a source signal, it aids in externalizing the source, although the degree of externalization seems situational and probably depends on the amount of low-frequency energy and the transient content of the source. The externalization of auditory images represents the most important difference between headphone and loudspeaker reproduction, and, in this sense, decorrelation helps to minimize the difference between the two modes of reproduction. Listeners also generally feel that the presence of decorrelated reverberation provides a more comfortable and relaxed listening experience, that is closer to listening in a natural environment.

Figure 11. Depiction of internalized and externalized sound images, resulting (respectively) from correlated and decorrelated sound sources.



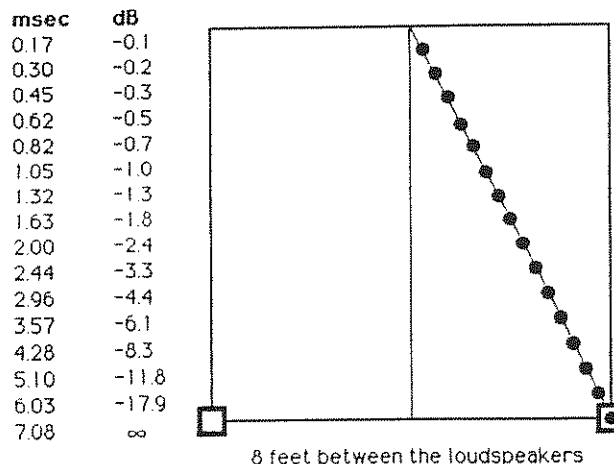
Effect No. 4: Reduction of "Image Shift" of Diffuse Sound Fields

When a time delay between identical sounds arriving from two loudspeakers is less than approximately 1.0 msec, listeners describe hearing a single sound image that is located between the loudspeakers. This is called "image shift" (Barron 1971). The sound image is shifted to the left or right, depending on whether the signal arrives first from the left or right loudspeaker, respectively. In stereo reproduction, image shift most typically occurs when listeners are located to either side of the center line that is equidistant from the loudspeakers.

Decorrelation causes a dramatic reduction in the image shift of the sound field. The effect is salient for listening positions at the extremes of the loudspeaker coverage. This was illustrated by Kendall, Wilde, and Martens (1989), who reported an experiment that used a combination of time delays and level differences typical of stereo reproduction in a small room. They compared the threshold for the collapse of the sound image into one loudspeaker in the case of correlated and decorrelated sound sources. This threshold is a function of both time delay and level difference. To relate these along a single continuum, they chose the sequence of simulated listening locations in a small room illustrated in Figure 12. The resulting time delay and level difference pairs are representative of those found in actual reproduction settings.

Figure 12. Simulated listener locations used by Kendall, Wilde, and Martens (1989). Each location is represented by a dot. For the sound arriving at the

listener location from the two loudspeakers, the time delay is given in msec and the intensity difference in dB.



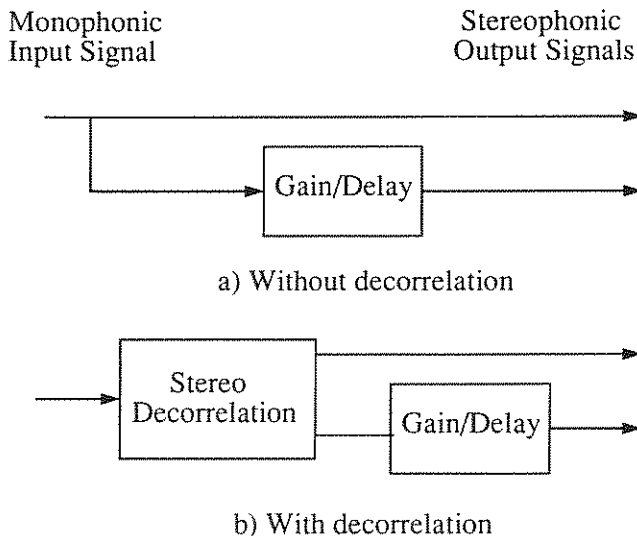
The experiment's stimuli were constructed from four monophonic sound sources chosen to vary greatly in their transient and sustained qualities, as described below.

1. "Snare": single snare drum stroke without reverb.
2. "Piano": single staccato chord in the middle register.
3. "Speech": recording of the sentence "I'm Batman."
4. "Quartet": single sustained chord extracted from a CD recording of Beethoven's *String Quartet* no. 12 in E-flat major/op. 127.

The complete set of stimuli included correlated and decorrelated stereo versions of these sources, to which a time delay and level difference had been added in one channel, as shown in Figure 13. (The stimuli were also equalized beforehand for correct overall level differences.) Subjects were seated in a small listening environment with sound absorption that removed early reflections from the walls near the loudspeakers (Kendall, Wilde, and Martens 1990). Subjects were asked whether the sound image was primarily located in one loudspeaker or not. The goal was to determine the threshold of level and time difference at which the sound source collapsed into one loudspeaker. The experiment was run as an adaptive two-alternative forced choice method. Two randomly interleaved stair-

Figure 13. Preparation of stimuli for the "image shift" and "precedence" experiments. Monophonic sources were "stereoized" with and without decorre-

lation. A time delay and level difference was added in one of the channels to create the stereo output signals.



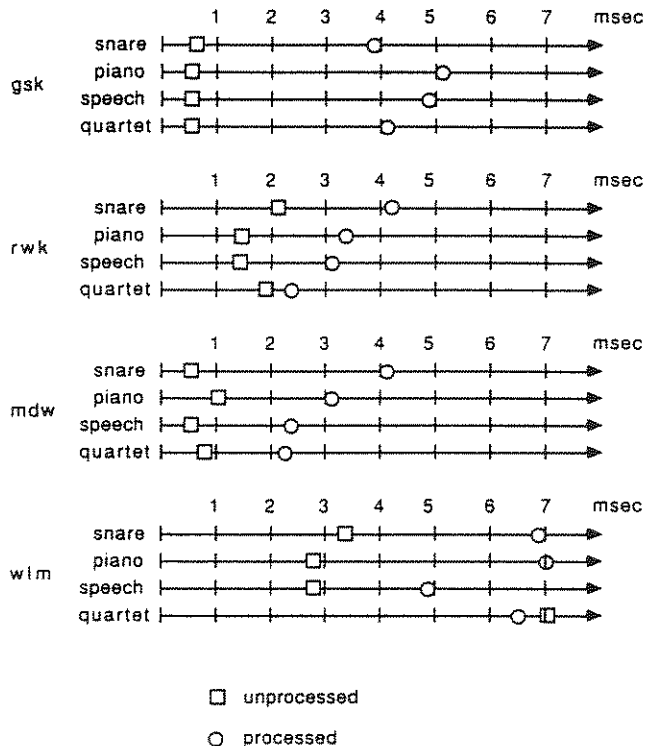
cases tracked independent estimates of the point at which the subject gave a "one" response 50 percent of the time.

The results are shown in Figure 14. The individual subjects (including the two researchers) apparently used different criteria to judge this threshold, but regardless of the criterion used, subjects judged the decorrelated stimuli as collapsing into a single loudspeaker much farther off to the side of the simulated listening room than the correlated stimuli.

Effect No. 5: Elimination of the Precedence Effect

Also called "the law of the first wavefront" and the "Haas effect," the "precedence effect" is the phenomenon in which a sound source in a natural environment is localized at the original source location while its reflected sound is ignored. The effect is particularly relevant for transient sounds. The precedence effect has typically been studied by delaying one sound source relative to another when reproduced with two loudspeakers. The effect became most familiar through the papers of Haas (1951) and Wallach, Newman, and Rosenzweig (1949). It is well reviewed by Gardener (1968), and our knowledge has grown with publications by Blauert (1971), Zurek (1980), and Lindemann (1986). A description of models of the precedence ef-

Figure 14. Thresholds for collapse of sound into one loudspeaker (Kendall, Wilde, and Martens 1989). Squares represent correlated stimuli, circles uncorrelated stimuli.



fect given by Rakerd and Hartmann (1985) states that it is a result of "a neural inhibition process which prevents the processing of binaural difference following an onset. There are indications that this inhibition is quite general, . . . there is some release from this binaural inhibition after approximately 10 ms and almost complete release within 50 ms (Zurek 1980)." (One manner in which the precedence effect appears to break down has been described by Clifton (1987): a sudden exchange of the directions of the leading and following signals results in a perception of two sources for a few seconds.)

While most discussions of the precedence effect relate it to the perception of reflected sound in natural environments, it is also a key factor in the perception of sound imagery over loudspeakers. In fact, the auditory system "interprets" loudspeaker reproduction in exactly the same way that it does environmental sound. This is most clearly illustrated when the auditory system inhibits the perception of sound arriving from a second, more dis-

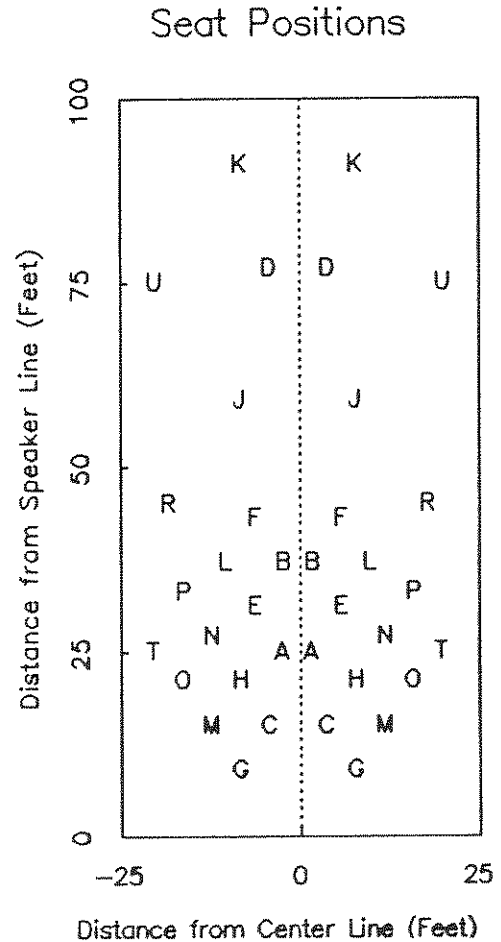
Figure 15. Simulated seating positions for experiment reported by Kendall, Wilde, and Martens (1989).

tant loudspeaker. In typical listening environments such as living rooms or theaters, most listeners are located nearer to one loudspeaker than to the other(s), and when the same sound material is reproduced by two or more loudspeakers, most listeners report that the sound images are entirely located in the nearest loudspeaker.

Our understanding of the conditions under which precedence operates is complicated by two factors. The first is that the precedence effect is more pronounced for transient sound sources such as struck or plucked musical instruments, than for continuous sound sources, such as blown or bowed musical instruments. The second complicating factor is that differences in arrival time are accompanied by differences in intensity, and the ratio between time delay and intensity difference varies tremendously across the potential listening positions in all reproduction environments.

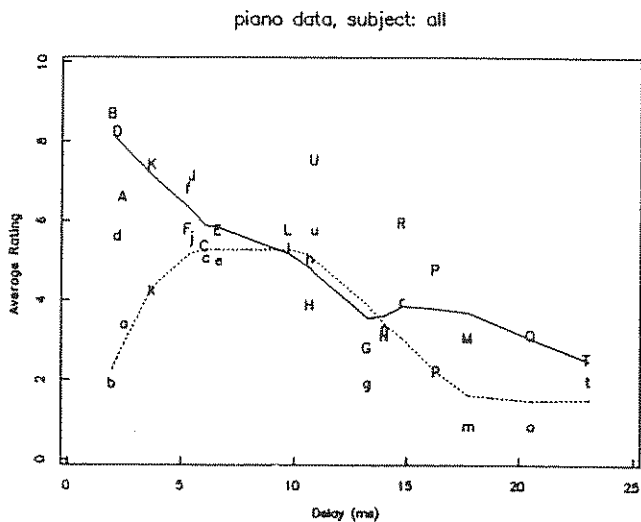
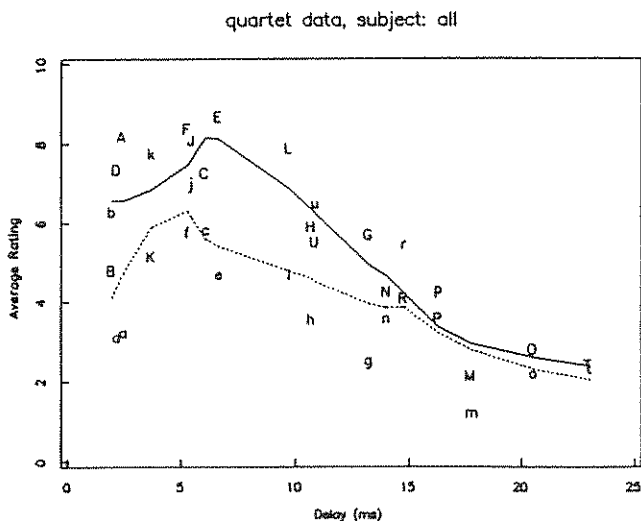
Kendall, Wilde, and Martens (1989) report a study that compares listeners' judgments of correlated and decorrelated sound sources under conditions that would generally invoke the precedence effect. The same stimuli described in Effect No. 4 above were prepared as before (see Figure 13), and subjects were seated in the same small listening environment without early reflections. In this experiment, the simulated listener locations were distributed widely across a 50 × 100 foot area, as shown in Figure 15. Each location represents a unique combination of delay time and level differences that could be anticipated to occur in a practical reproduction setting. Time delays range from 2 to 23 msec. Seating locations are identified with letters moving alphabetically from the center of the room toward the outside wall. (Some letters are missing because the seating locations associated with those letters were dropped when the number of stimuli prepared for the experiment was trimmed down.) Subjects were asked to rate each sound image on a 10-point scale. A rating of 0 represented an image that was collapsed completely into a single loudspeaker; a rating of 10 represented a split image that was divided between the loudspeakers, or a single image that was located between the loudspeakers (the latter happening when the time delay was short).

Figures 16a and 16b show averaged ratings from



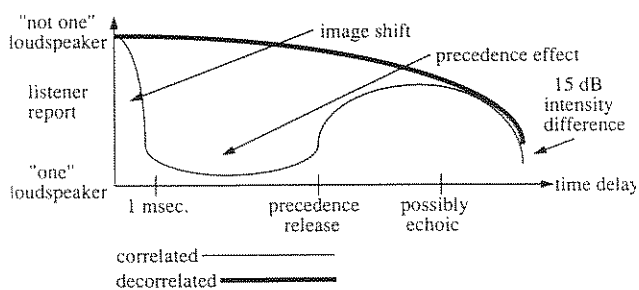
all subjects for the quartet and the piano stimuli, respectively. The lower-case and upper-case letters are associated with each seating location shown in Figure 15, and represent responses for correlated and uncorrelated stimuli, respectively. The horizontal axis shows the time delay associated with the seating locations. These responses are somewhat scattered due to the variations in intensity difference at seating locations with the same time delay. Trends clearly emerge in the averaged rating for each delay. The broken and solid lines represent averaged ratings for correlated and decorrelated stimuli, respectively. In the range of short delays (from 2 msec up to 6 or 7 msec), precedence clearly dominates the correlated, but not the decorrelated, stimuli. Correlated stimuli are heard mostly in one

Figure 16. Averaged ratings from all listeners for piano (a) and quartet (b) (Kendall, Wilde, and Martens 1989). The broken and solid lines represent averaged ratings for correlated and decorrelated stimuli, respectively.



loudspeaker, while the decorrelated stimuli are heard in two. Ratings vary with each sound source. Precedence affects the piano stimuli somewhat more than the quartet, most likely because it is more transient. Above 10 msec, the precedence effect is "released," and both the correlated and decorrelated stimuli begin to collapse toward one loudspeaker when intensity differences approach 15 dB.

Figure 17. A summary depiction of listeners' subjective impressions of both correlated and decorrelated sound sources across a wide range of delays (Kendall, Wilde, and Martens 1989).



In large-space reproduction, the use of decorrelation inhibits the collapse of sound imagery that is due to the listener's seating location. It also helps to smooth out the differences in imagery between transient and non-transient sounds. When there are more than two loudspeakers, multiple channels of decorrelated sound can be created, following the example of Figure 5. This kind of multi-channel decorrelation helps to maintain the listener's awareness of all the loudspeakers in the reproduction space and stabilize spatial imagery.

Conclusion

Figure 17 provides a summary depiction of listeners' subjective impressions of spatial imagery for both correlated and decorrelated sound sources. The vertical axis represents the listeners' subjective judgment of whether the sound image is located primarily in one loudspeaker or not. The horizontal axis represents the difference in arrival time between the nearer and farther loudspeaker. (Intensity differences generally increase as time differences increase, but are not represented in this figure.) The spatial imagery of correlated sound sources varies tremendously with time delay. When the time delay is less than approximately 1.0 msec, listeners describe hearing a single sound image that is located between the loudspeakers ("image shift") and when the time delay is greater than approximately 1.0 msec, listeners describe hearing a single sound image that is located at the closer loudspeaker ("precedence effect"). At some higher time delay, the precedence effect is released, and the sound

will be heard in both loudspeakers. (The exact delay at which the precedence effect is released depends upon the transient qualities of the particular sound source.) When the loudspeakers are separated by a sufficiently great distance, listeners report that the delayed sound is like an echo. As the time delay further increases, the intensity difference increases until at approximately 15 dB listeners again report that the sound image is located in one loudspeaker. These sorts of radical changes in sound imagery show up vividly when audio material is moved from one reproduction setting to another, for example, from the studio to the concert hall. As shown in Figure 17, decorrelation minimizes these radical changes (and promotes spatial imagery that will remain invariant in divergent reproduction settings). The imagery of decorrelated sounds varies little throughout the range of delays typically associated with "image shift" and the "precedence effect." It also provides for externalization in headphone reproduction, which again is a stabilizing influence on spatial imagery.

An understanding of the effects of decorrelation provides an additional dimension to the work of sound artists and audio professionals. It offers nuances to imagery and expands the range of aesthetic possibilities. It improves the consistency of sound imagery in the wide variety of reproduction settings encountered every day, so that the artist's intentions for spatial imagery are much more likely to be communicated to the audience.

Acknowledgments

Deepest appreciation and thanks to William Martens and Martin Wilde for their friendship and support during years of work together. Thanks to Matt Moller for his help in implementing new software tools for decorrelation. Thanks also to Doug Keislar for his careful and thoughtful comments on the manuscript. Portions of this paper are taken from Kendall (1994). Commercial application of this work is covered by US patent no. 5,235,646, assigned to Northwestern University.

References

- Ando, Y. 1977. "Subjective Preference in Relation to Objective Parameters of Music Sound Fields with a Single Echo." *Journal of the Acoustical Society of America* 62:1436-1441.
- Augspurger, G., et al. 1989. "Use of Stereo Synthesis to Reduce Subjective/Objective Interference Effects: The Perception of Comb Filtering, Part II." Preprint 2862; the 87th Convention of the Audio Engineering Society, October 1989, New York.
- Barron, M. 1971. "The Subjective Effects of First Reflections in Concert Halls—The Need For Lateral Reflections." *Journal of Sound and Vibration* 15:475-494.
- Blauert, J. 1971. "Localization and the Law of the First Wavefront in the Median Plane." *Journal of the Acoustical Society of America* 50:466-470.
- Blauert, J., and W. Lindemann. 1986. "Spatial Mapping of Intracranial Auditory Events for Various Degrees of Interaural Coherence." *Journal of the Acoustical Society of America* 79:806-813.
- Chowning, J. M. 1971. "The Simulation of Moving Sound Sources." *Journal of the Audio Engineering Society* 19:2-6. Reprinted in *Computer Music Journal* 1(3):48-52.
- Clifton, R. K. 1987. "Breakdown of Echo Suppression in the Precedence Effect." *Journal of the Acoustical Society of America* 82:1834-1835.
- Durlach, N. I., et al. "On the Externalization of Auditory Images." *Presence* 1(2):251-257.
- Gardener, M. B. 1968. "Historical Background of the Haas and/or Precedence Effect." *Journal of the Acoustical Society of America* 43:1243-1248.
- Haas, H. 1951. "Über den Einfluss eines Einfachechos auf die Hörsamkeit von Sprache." *Acustica* 1:49-58.
- Kendall, G. 1994. "The Effects of Multi-Channel Signal Decorrelation in Audio Reproduction." *Proceedings of the 1994 International Computer Music Conference*. San Francisco: International Computer Music Association.
- Kendall, G. S. 1989. Unpublished personal lab notebook.
- 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*. San Francisco: International Computer Music Association.
- Kendall, G. S., M. D. Wilde, and W. L. Martens. 1989. "Production and Reproduction of Three-Dimensional Sound." Paper presented to the Audio Engineering Society 87th Convention, New York.

- Kendall, G. S., M. D. Wilde, and W. L. Martens. 1990. "A Spatial Sound Processor for Loudspeaker and Headphone Reproduction." *Proceedings of the Audio Engineering Society 8th International Conference*. New York: Audio Engineering Society.
- Kurozumi, K., and K. Ohgushi. 1983. "The Relationship between the Cross-Correlation Coefficient of Two-Channel Acoustic Signals and Sound Image Quality." *Journal of the Acoustical Society of America* 74:1728–1733.
- Lindemann, W. 1986. "Extension of a Binaural Cross-Correlation Model by Contralateral Inhibition, II. The law of the First Wave Front." *Journal of the Acoustical Society of America* 80:1623–1630.
- Plomp, R., and H. J. M. Steeneken. 1969. "Effect of Phase on the Timbre of Complex Tones." *Journal of the Acoustical Society of America* 46:409–421.
- Pollack, I., and W. J. Tritpoe. 1959. "Binaural Listening and Interaural Cross Correlation." *Journal of the Acoustical Society of America* 31:1250–1252.
- Rakerd, B., and W. M. Hartmann. 1985. "Localization of Sound in Rooms, II: The Effects of a Single Reflecting Surface." *Journal of the Acoustical Society of America* 78(2):524–533.
- Schroeder, M. R. 1984. *Number Theory in Science and Communication*. Berlin: Springer-Verlag.
- Schroeder, M. R., D. Gottlob, and K. F. Siebrasse. 1974. "Comparative Study of European Concert Halls: Correlation of Subjective Preference with Geometric and Acoustic Parameters." *Journal of the Acoustical Society of America* 56:1195–1201.
- Wallach, H., E. B. Newman, and M. R. Rosenzweig. 1949. "The Precedence Effect in Sound Localization." *The American Journal of Psychology* 62:315–336.
- Wilde, M. D. 1989. "The Psychoacoustical Effects of Interaural Cross-Correlation." master's thesis, Northwestern University.
- Wilde, M., G. Kendall, and W. Martens. 1990. "Method for Controlling the Width and Distance of an Acoustical Image." Unpublished patent application assigned to the former Auris Corporation.
- Wilde, M., W. Martens, and G. Kendall. 1989. "Apparatus for Creating Decorrelated Audio Output Signals with a Specified Cross Correlation Coefficient." Unpublished patent disclosure to Northwestern University.
- Zurek, P. M. 1980. "The Precedence Effect and Its Possible Role in the Avoidance of Interaural Ambiguities." *Journal of the Acoustical Society of America* 67:952–964.

Appendix: Controlling Image Distance

The perceived distance of a sound image reproduced over stereo loudspeakers can be affected in several ways. Chowning (1971) describes a method for controlling the ratio of direct to reverberant sound that creates distance illusions akin to those in concert halls. Kendall and Martens (1984) describe the use of simulated early reflections without reverberation to create vivid distance cues typical of smaller rooms. Both of these techniques create illusions of image distance that lie beyond the loudspeakers (except in special circumstances, in which the motion of a sound image leads the listener to infer that it must have passed closer).

As a by-product of the FIR direct filter design techniques described above, a discovery was made affecting how the distance of a sound image between the loudspeakers and the listener could be controlled (Kendall 1989; Wilde, Kendall, and Martens 1990). As described above, the work of Kurozumi and Ohgushi (1983) as well as Wilde (1989) demonstrates that the distance of a sound image from the listener will vary directly with the value of the correlation measure. A correlation measure near 1 is associated with sound images in the plane of the loudspeakers, a correlation measure near 0 is associated with sound images between the loudspeakers and the listener, and a correlation near -1 produces sound images near the listener's head.

These changes in image distance produced by decorrelation are also accompanied by changes in image width. This dual effect begs the question (not answered above) of which acoustic factors affect image distance, and which affect image width. Width is clearly associated with the randomization of inter-channel phase relationships, while distance is associated with the shift from a correlation measure of $+1$ to a correlation measure of -1 . The difference between $+1$ and -1 correlation is easily demonstrated by flipping stereo signals "in phase" and "out of phase." The challenge of finding a method of continuous transformation from "in phase" to "out of phase" signals was solved by creating a set of filters with a constant phase offset.

Figure 18 illustrates the filter-design technique

Figure 18. The computation of coefficients for a filter with a constant phase offset, $\Delta\theta$. The left channel is unprocessed, and the right-channel fil-

ter creates the inter-channel phase difference. The correlation measure, Ω , varies from +1 to 0 to -1 as $\Delta\theta$ varies from 0 to $\pi/2$ to π .

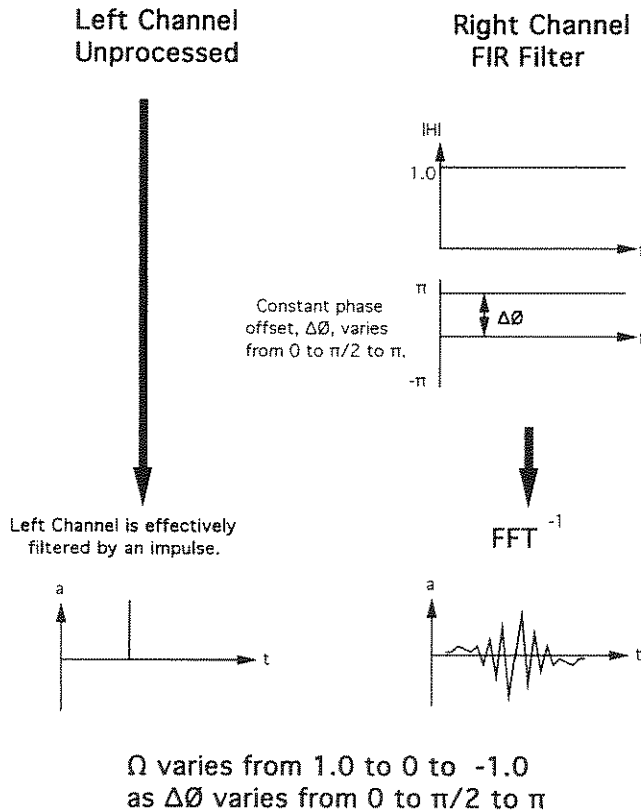
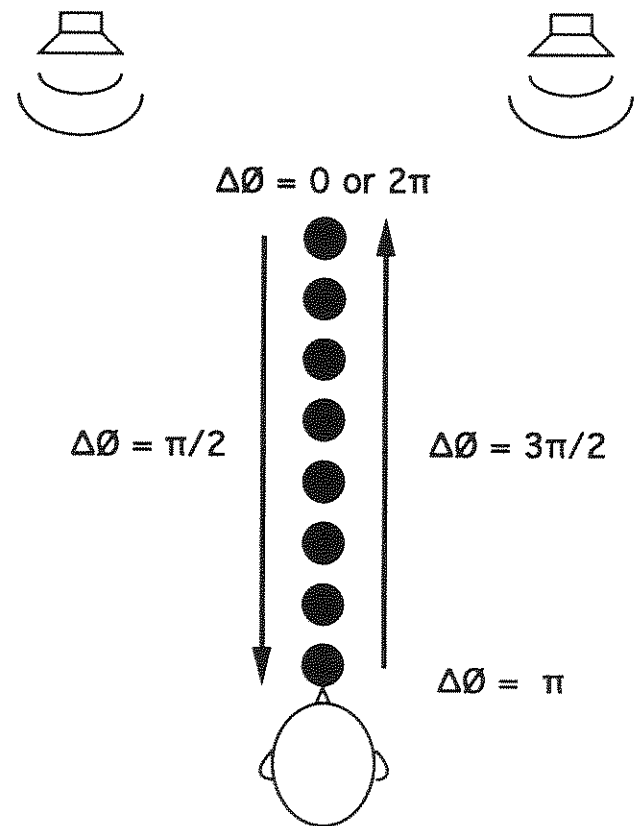


Figure 19. Depiction of changes in image distance reported by listeners as the constant phase offset, $\Delta\theta$, varies from 0 to 2π .



Sound Image Movement

that produces correlation measures from +1 to -1 without randomizing phase. First, the left channel of the input signal is left unprocessed (or, one could say, is in effect filtered by an impulse). Second, the right channel is filtered by an all-pass FIR filter with constant phase. The filter is designed by specifying a constant magnitude of 1.0 and a constant phase with a value within the range from 0 to π . The frequency-domain specification is converted to filter coefficients via the IFFT. The correlation measure of the resulting FIR filter coefficients varies from +1.0 to 0 to -1.0 as the constant phase offset, $\Delta\theta$, varies from 0 to $\pi/2$ to π .

When presented with an audio demonstration of sounds with constant phase offset varying from 0 to 2π , listeners report a series of sound images that move from the plane of the loudspeakers to the lis-

tener's head and back again, as depicted in Figure 19. This technique appears to be a simple method of capturing the kind of interaural phase changes that occur when a sound source is close to the head. It therefore suffers the same type of dependence on listener location that is typical of loudspeaker reproduction of head-related transfer functions. It is quite effective for near-field monitoring, such as occurs with properly arranged home stereo systems, stereo television, or stereo computer monitors, but rather useless for large-space reproduction in concert halls or theaters.