Fixing the Phantom Center:
Diffusing Acoustical Crosstalk

Earl Vickers[1]

[1] STMicroelectronics, Inc., Santa Clara, CA 95054, USA
earl.vickers@st.com

ABSTRACT

When two loudspeakers play the same signal, a "phantom center" image is produced between the speakers. However, this image differs from one produced by a real center speaker. In particular, acoustical crosstalk produces a comb-filtering effect, with cancellations that may be in the frequency range needed for the intelligibility of speech. We present a method for using phase decorrelation to fill in these gaps and produce a flatter magnitude response, reducing coloration and potentially enhancing dialogue clarity. This method also improves headphone compatibility, and it reduces the tendency of the phantom image to move toward the nearest speaker.

0.   INTRODUCTION

In this article, we will explore some of the issues of the "phantom center" image, with an emphasis on the comb-like frequency responses produced when correlated signals are played through a pair of loudspeakers. We will review some earlier solutions to the problem and present a new solution.

0.1.   Intensity Stereo and Phantom Images

At this point, it is easy to forget that stereo is every bit as much an illusion as "virtual surround" and similar technologies. It now seems obvious that a sound image can be panned across a phantom soundstage simply by varying the relative amplitude of two speakers, but the way in which this illusion works is still somewhat surprising.

In 1931, Blumlein [1] showed that at low frequencies, amplitude differences between a pair of loudspeakers produce phase differences between the ears, thus mimicking the phase differences that occur in normal listening and producing phantom images in the desired locations. (Conversely, phase differences between the speaker signals produce level differences at the ears [2].) These effects occur as a result of summing of the acoustical pressure phasors from the two speakers at the two ears [3].

1. PROBLEMS WITH THE PHANTOM CENTER

Unfortunately, the phantom center illusion is not perfect. There are a number of problems, including issues with localization, mono-compatibility, headphone compatibility, and frequency response.

We will be primarily concerned with the frequency response, but we will briefly review some of the other issues.

1.1. Phantom Center Moves toward the Nearest Speaker

First of all, the precedence effect (also called the "law of the first wavefront" or the "Haas effect," actually discovered by Joseph Henry [2][4]) says that identical transient sounds will appear to come from the closest source, causing the image to follow you toward the nearest speaker when you change positions. When viewing audio/visual content, this can cause a localization mismatch between sound and picture.

1.2. Mono-Compatibility

In 1968, Eargle addressed the issue that while two equal signals add electrically to give a gain of 6 dB, they add acoustically for a gain of 3 dB, because the effective reverberant sound pressure of two speakers in a reverberant sound field is

$$P_E = \sqrt{P_L{}^2 + P_R{}^2} \ , \qquad\qquad (1)$$

where P_E is the effective reverberant sound pressure, and P_L and P_R are the sound pressures from the left and right speakers, respectively. As a result, the center balance changes when downmixing to one channel; there is a 3 dB "center build-up" when a stereo program is played back in mono (for example, over AM radio) [5]. In addition, downmixing to mono reduces the reverberant-to-direct ratio by 3 dB [6].

Eargle proposed a "quadrature split" method, in which the center channel is split into left and right components such that both halves are of equal magnitude and 90° out of phase. In this way, summing the left and right channels will always yield an increase of 3 dB, regardless of whether they are combined acoustically or electrically [5].

However, the quadrature phase relationship may affect the localization of the phantom center [5]. Bauer stated that a 90° phase relationship would spread the phantom center across the entire soundstage between the speakers, due to conflicting phase and magnitude cues [7]. This spreading effect is best perceived in a relatively "dead" room, with the listener equidistant from the speakers [5]. Furthermore, the conflict between the cues, and hence the spreading effect, should occur primarily at low frequencies (below around 1 kHz), since phase cues become ambiguous at high frequencies.

1.3. Speaker / Headphone Compatibility

When listening to stereo audio over headphones, the sound generally appears to come from inside the head, and the phantom center is localized in the center of the head. Many listeners find this effect uncomfortable and fatiguing. Bauer presented a method of spreading the headphone image using a simple 90° phase-splitting network [8].

1.4. Comb Filtering Notches

Acoustical crosstalk occurs when the same signal from a pair of speakers reaches the ear at slightly different times. While the resulting phase changes facilitate the intensity stereo illusion at low frequencies, they also create a comb-filtering effect, with a series of magnitude notches across the frequency spectrum, as illustrated in Figure 1. This coloration not only implies that the phantom center image will always sound somewhat different from a real center speaker, but it also appears to reduce the intelligibility of speech [9].
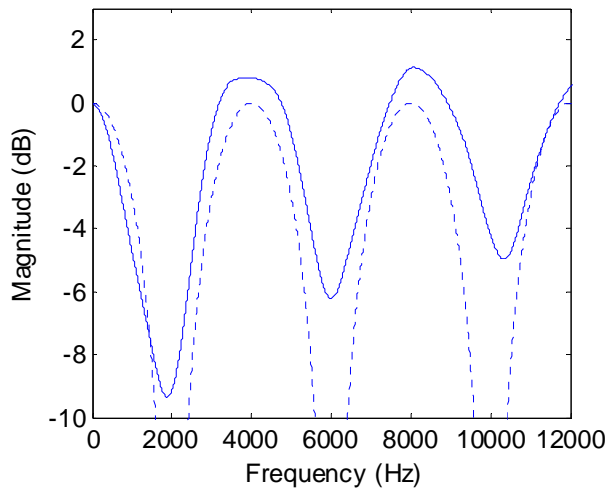
Figure 1. Simulated comb filtering notches from interaural crosstalk, at one ear. Spherical head model (solid line) after Duda et al [10]; simple time delay model (dashed line). The comb filtering is less severe with the spherical model, due to head shadowing and diffraction.

Holman reported that the phantom center image seemed less clear, compared to the sound from a real center speaker, and that more work was required in order to understand dialogue [6]. Shirley et al showed a small but statistically significant decrease in speech intelligibility when using the phantom center instead of a real center speaker [9]. They suggested that this difference was caused by the phantom center's acoustical crosstalk and resulting magnitude dips. We will focus primarily on the lowest-frequency notch, which is centered around 2 kHz for the standard speaker angles of ±30° and at higher frequencies for narrower angles. This is not only the deepest notch, but it also occurs in a frequency band that is important to speech recognition [9].

Note that if we regard the speech (which is typically panned toward the center) as "signal" and the uncorrelated components of music and sound effects as "noise," the signal-to-noise ratio will be worse at the frequencies surrounding the notch, because the uncorrelated ("side") components are not subject to the cancellation. As a result, the energy of the side images is likely to mask that of the phantom center near the notch frequency.

1.4.1. Not Noticing What We Don't Notice

There are a number of acoustical phenomena that are rarely noticed consciously, for various reasons. These include individual harmonics (which perceptually fuse with the fundamental), early reflections (which perceptually fuse with the direct sound), late reverberation (which generally goes unnoticed unless it exceeds a certain level), sounds masked due to spectral or temporal proximity to other sounds [11], and spectral notches produced by comb filtering.

It is rather surprising that we fail to notice the coloration and missing spectral information caused by the deep notches of acoustical crosstalk. If we were to find such an uneven frequency response in a speaker we had just bought, we would likely return it for a refund. And yet, this defective response is universally provided as a free bonus with the purchase of a second speaker.

Discussing this topic, David Clark stated:

"Clearly the 'phantom' center is an entirely different listening experience than pure left or pure right. One might ask if stereo is deeply flawed as [a] sound reproduction technique or if interference notches should simply be ignored [12]."

When the listener is not equidistant from the speakers, the comb filter peaks and nulls will be in different frequency locations at the two ears, as shown in the simple time-delay simulation of Figure 2. (In real life, the responses would be significantly more complicated, especially at high frequencies, due to head-shadowing, pinna reflections and other effects.)

While the comb filters of Figure 2 are not complementary, the resulting interaural intensity differences somewhat resemble those caused created by typical "pseudo-stereo" algorithms such as the Lauridsen decorrelator (complementary comb filters) [13], to the extent that it's surprising that the phantom center does not spread to cover the entire range between the speakers. While the phantom center image is sometimes regarded as slightly broader or less spatially constricted than that of a real center speaker [14], it is nevertheless readily localized to a small area.
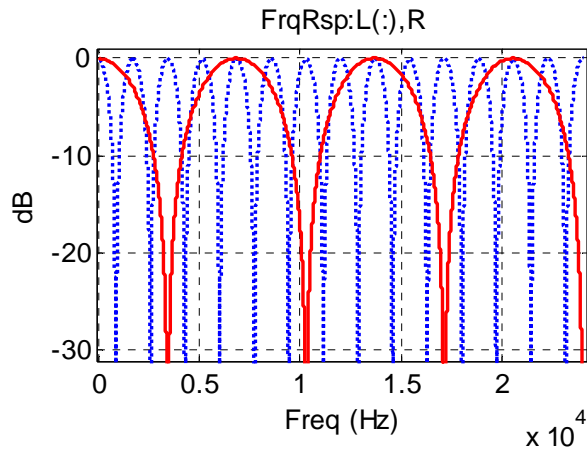
FrqRsp:L(:),R

Figure 2. Magnitude response of a simple time-delay
model of asymmetrical phantom mono
acoustical crosstalk at left ear (dotted line)
and right ear (solid line).

We can postulate a number of possible reasons for the
fact that we seldom notice these comb filter notches and
pseudo-stereo effects when listening to audio over two
speakers. It is known that narrow spectral notches are
far less obvious than narrow peaks [15]. Also, these
notches occur at mid to high frequencies, while people
tend to pay more conscious attention to the lower
frequency ranges which contain the fundamental pitches
of voice and music [16]. When audio content includes
correlated (center) and decorrelated (side) information,
only the center content is subject to the comb filtering,
reducing the salience of the notches. Theile's
"association model" suggests that these spectral
variations are not directly perceived, because the brain
associates them with the binaural information used for
localization in normal listening [14][17]. Furthermore,
the perceived sound color may be a function of the
average spectrum of the two ear signals; this tends to
smooth the spectrum because the signals at the two ears
generally have different peaks and notches [18]. In
addition, for frequencies above 2 kHz, the comb
filtering is reduced by head shadowing [18].

1.4.2. Comb Filter Notches Smoothed by Room Reflections

Yet another reason we do not consciously notice these
comb filter notches is that room reflections and
reverberation from all directions, while creating new
cancellations of their own, nevertheless tend to smooth
out the magnitude responses, filling in some of the
missing information.

As mentioned, early reflections perceptually fuse with
the direct sound, adding more useful energy and
increasing the intelligibility of speech [19]. The ability
of room reflections to fill in the phantom center notches
may be another reason for the improved intelligibility.

On the subject of comb filtering and room reflections,
Clark stated:

- "Response notches are annoying if not filled in by
  reflections."

- "Response notches are almost inaudible if the notches
  are filled in by reflections within 10 ms."

- "Stereo reproduction depends on room reflections."
  [!]

- "Very dead rooms leave audible comb filtering."

- "Perhaps as the interference pattern of the stereo
  speakers becomes sufficiently dense it is perceived as
  adding to spaciousness. Listeners frequently
  commented that two speaker mono sounds fuller,
  more solid or has depth." [12]

Finally, Clark states:

- "Any sacrifice in timing or phase shift should be
  made if it will produce a better frequency response."
  [12]

We will return to this latter idea.

2.   PREVIOUS SOLUTIONS

2.1.   Physical Center Speaker

One solution to the problems of phantom center images
is simply to add a real center speaker. This was the
approach taken by Harvey Fletcher of Bell Labs, at
roughly the same time (ca. 1933) that Blumlein was
inventing two-channel intensity stereo. This approach
had the advantage of providing a stable center image.
When Walt Disney produced "Fantasia" in the late
1930s, the dialogue was assigned to a center channel,
behind the screen [20], thus avoiding the problem of
off-center listeners localizing the dialogue toward the
nearest side speaker.

For reasons of cost and space, many consumer audio and television systems do not include a center speaker. Therefore, we would like to find an approach that works over two speakers.

2.2.   Crosstalk Cancellation

A now-obvious solution to the problem of acoustic crosstalk is to cancel it before it happens [21]. Unfortunately, at mid and high frequencies, this is effective only within a relatively small "sweet spot," which limits the usefulness of this technique for typical television viewing and other situations involving multiple listeners in arbitrary positions.

2.3.   Equalization

Another way to fix the non-flat magnitude response caused by acoustic crosstalk is to apply inverse filters to the left and right signals [22].

However, the frequencies of the comb filter notches vary greatly depending on the relative positions of the speakers and listener. For example, the cancellation frequencies increase as the angle subtended by the speakers becomes narrower (such as when the listener moves further back) [9]. In addition, as the listener moves to the side and is no longer equidistant from the speakers, the notches move closer together and become different for each ear [12]. Without a good estimate of the relative positions, it would be impossible to accurately equalize the effects of the crosstalk.

3.   IMPROVING THE PHANTOM CENTER MAGNITUDE RESPONSE

Figure 3 is a simplified top-down view of an asymmetrical listening environment. The loudspeakers are shown at the upper left and right, the listener is represented by the circle at the bottom of the diagram, and the speaker-to-ear paths are shown by the diagonal lines.

Even in the unlikely case that the center of the listener's head is located exactly along the plane of symmetry between the speakers, the ears will not be, assuming the listener is facing forward. At each ear, the dual mono signal will be received from the two sources with different time delays.
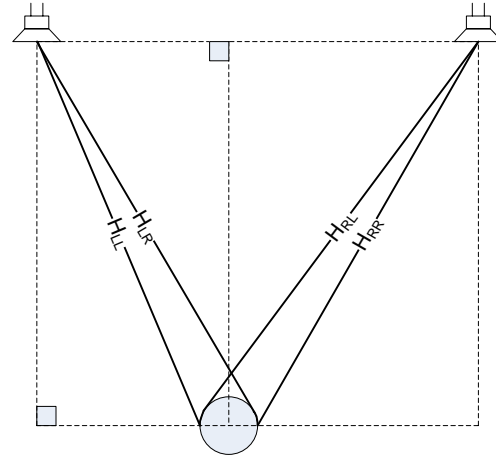

Figure 3. Simplified, possibly asymmetrical listening geometry.

The crosstalk can be modeled by a system like that of Figure 4, in which transfer functions H_{LL} and H_{RR} represent the ipsilateral paths from left speaker to left ear and from right speaker to right ear, respectively, and H_{LR} and H_{RL} represent the contralateral paths from left speaker to right ear and from right speaker to left ear, respectively.
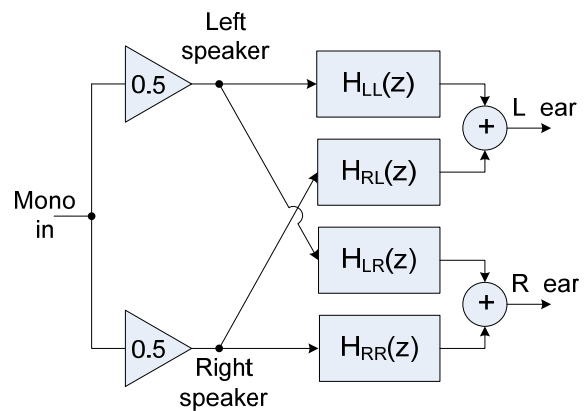

Figure 4. Simplified phantom mono acoustical crosstalk model.

The transfer functions from the mono input to the two ears are given by

$$H_L(z) = 0.5[H_{LL}(z) + H_{RL}(z)], \text{ and} \qquad (2)$$

$$H_R(z) = 0.5[H_{LR}(z) + H_{RR}(z)]. \qquad (3)$$

For the moment, we will ignore details like head-shadowing, and focus only on the phase cancellations. In this case,

$$H_L(z) = 0.5\left[z^{-LL} + z^{-RL}\right], \text{ and} \qquad (4)$$

$$H_R(z) = 0.5\left[z^{-LR} + z^{-RR}\right], \qquad (5)$$

where LL and RR are the ipsilateral delays and LR and RL are the contralateral delays, measured in samples. A typical magnitude response was shown as the dotted line in Figure 1.

Whenever the acoustical delays from the left and right speakers to a single point (such as one ear) are unequal, there will be a series of frequencies at which the signals will be 180° out of phase. Even if we manage to add the right amount of electrical delay to equalize the acoustical delays from the left and right speakers to the left ear, the total delays will then be unequal at the right ear.

However, for intelligibility of speech consonants, it is not necessary to have a flat magnitude response at every frequency, due to the ear's "auditory filters." The ear assigns the same perceived loudness to narrow-band noise sources, regardless of the noise bandwidth, so long as that bandwidth is less than the "critical bandwidth." Thus, even if there are cancellations within a given critical band, what matters is the total noise power within that band [11].

This simplifies the problem considerably. Our proposed solution involves decorrelating the phase differences between channels, within each critical band, effectively randomizing the cancellations and reducing their perceived effect.

3.1. Decorrelation Methods

There are many methods of generating diffused, decorrelated signals, including Feedback Delay Network (FDN) reverbs [23][24] and convolution with time-limited white noise [25] and "velvet noise" [26]. Liu and Smith explored a number of ways of generating perceptually similar copies from a mono sound [27]. Kendall presented an excellent overview of decorrelation methods and the impact of decorrelation on spatial imagery and coloration [28].

Bouéri and Kyriakakis decorrelated audio signals by applying random time shifts to each critical band. They reported a wider, more spacious image and an elimination of the Haas effect, with a slight change in timbre during multiple-speaker playback [29].

In the context of controlling the effects of crosstalk, Augspurger et al [30] used comb filtering to fight comb filtering, based on the hypothesis that dense "pre-combing" might reduce the phase cancellations caused by multiple arrival times. They tested four different filters: comb filters using constant time delay, Hilbert transform using constant phase shift, simple dual mono, and a commercial stereo synthesizer which produced comb filtering on a log frequency scale. Tests were conducted in the 18,000-seat Hollywood Bowl amphitheater, which had problems with cancellations between multiple horns and drivers, as well as "swishiness" (phasiness?) in the mid- and high-frequencies above about 500 Hz.

Their experiments showed that pre-combing helped scramble the periodic nulls that would otherwise occur. They wrote, "This property is useful for restoring most of the harmonic spectrum which tends to be lost when both the fundamental and its overtones are cancelled due to the presence of integer multiples of acoustical half wave length between the spaced loudspeakers. The solution and the problem are ironically the same."

The experiments resulted in significant improvement in subjective impressions of clarity and coverage. However, there were some issues with timbre change, and the researchers did not test how the system might affect spatialization of stereo inputs. [30]

3.2. Schroeder's "Improved Quasi-Stereophony"

We would like to find an inexpensive method of decorrelating the phases between the two speakers, while allowing the output of each speaker to be approximately allpass (in case a listener is seated closer to one speaker than another).

In 1961, Schroeder showed that a pseudo-stereo effect could be produced by creating phase differences using a pair of allpass filters, as shown in Figure 5 [31]. In this circuit, each output is allpass if gain g equals $\pm\sqrt{0.5}$, but the two impulse responses are different due to the sign difference between the two multipliers. Therefore, the phase responses are different, producing envelope delay

differences as a function of frequency. Despite the fact that both outputs have the same (flat) magnitude response, the circuit can create pronounced spatial effects. This finding led Schroeder to revise his earlier view that differences in magnitude response were necessary in order to create a pseudo-stereo effect [13].
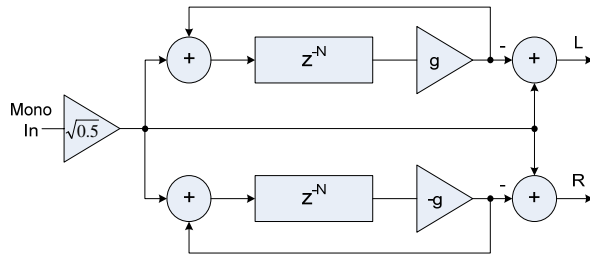

Figure 5. Schroeder's "improved quasi-stereo" filter with flat magnitude response, after [31].

3.3. Allowing Arbitrary Gains

The outputs of the above system are not allpass if gain g does not equal $\pm\sqrt{0.5}$, so we will substitute the system of Figure 6, which allows more flexibility.
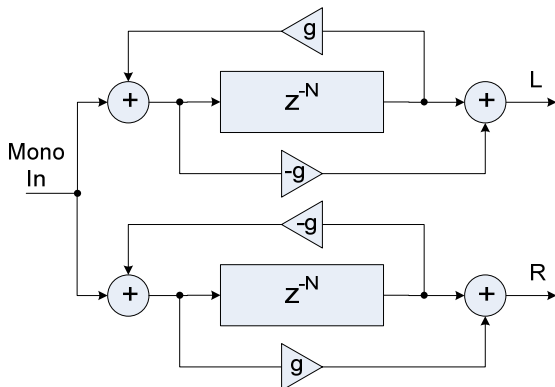

Figure 6. Phase diffusion based on Schroeder quasi-stereo, with arbitrary g.

This system has allpass transfer functions $A_L(z)$ and $A_R(z)$, as follows:

$$A_L(z) = \frac{-g + z^{-N}}{1 - gz^{-N}}, \text{ and} \tag{6}$$

$$A_R(z) = \frac{g + z^{-N}}{1 + gz^{-N}}. \tag{7}$$

3.4. Impulse Responses

The impulse responses are shown in Figure 7 for a gain of $g = 0.414$. Note that the decays are basically exponential (with the exception of the first pulse), and alternate pulses are opposite in sign at the two outputs.
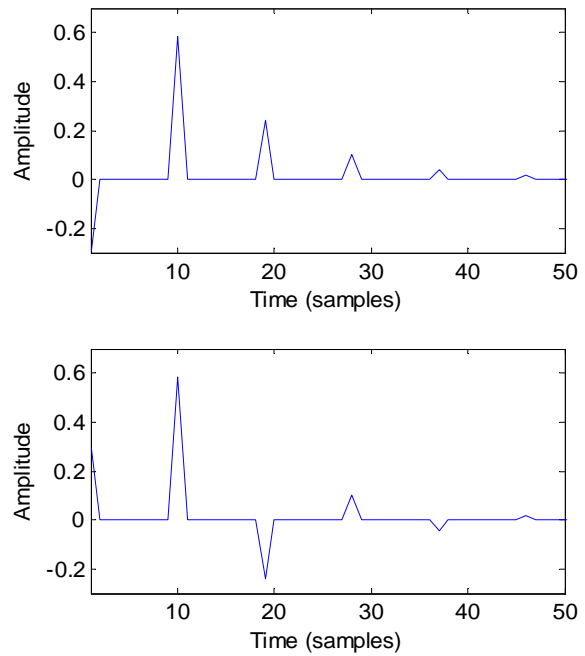

Figure 7. Left (top) and right (bottom) impulse responses of the filters shown in Figure 6.

The impulse responses are power-complementary (since both are allpass) but they are not in fact allpass-complementary (i.e., the phasor sum of the two impulse response outputs does not have constant amplitude), so they are not exactly mono-compatible. However, this system would normally be used for playback, not for encoding or signal transmission, so there would be no need to mix the output back to mono.

3.5. Magnitude and Phase Responses

The magnitude responses of the two outputs are flat, as mentioned, and the phase responses are shown in Figure 8 for a delay of $N = 9$ samples.
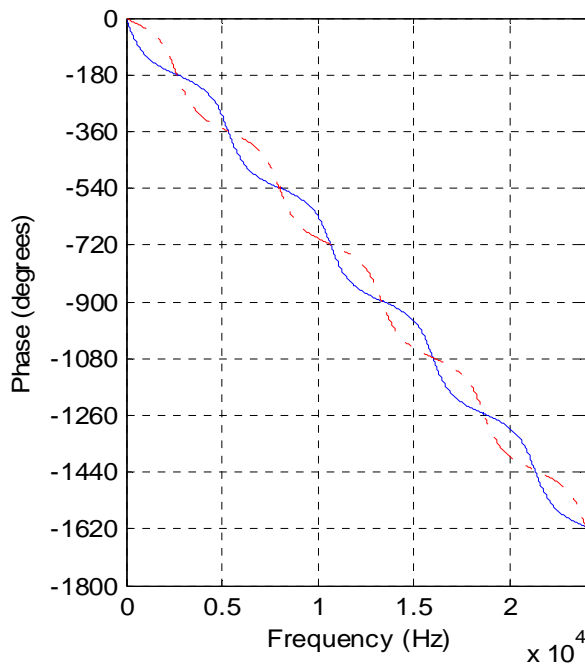
Figure 8. Left (solid) and right (dash-dot)
phase responses for N = 9 bands.

Note that the left and right phases are interleaved, so
that the left speaker leads at some frequencies and lags
at others. The number of alternating "bands"
corresponds to the delay length N. (If more than two
speaker signals are needed, additional decorrelated
outputs can be created by using different values of N.)

The difference between the left and right phases, as a
function of frequency, is shown in Figure 9. The phase
differences are somewhat proportional to the allpass
gain g. The value $g = 0.414$ has been selected
empirically so that the resulting phase differences cover
the range between approximately ±90°. The phase
difference increases monotonically, though not exactly
linearly, from 0° toward ±180° as g increases from 0
and approaches 1.0.

Møller et al [32] provide a good overview of the audible
effects of allpass filters. Ringing or "pitchiness" can be
apparent from second-order allpass filters with high Q,
but for typical signals these effects are likely to be
masked by the signals themselves [32]. With first-order
allpass filters, noticeable ringing can be caused by large
gain values, such as $g = 0.95$, but no such effects were
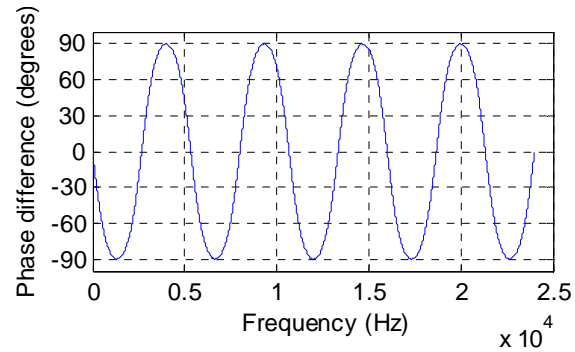noticed with $g = 0.414$.

Figure 9. Phase difference between the left and right
channels, as a function of frequency.

3.6.   Selecting the Delay Length

We would like to choose the delay N to be long enough
that we have at least one or two alternating phase bands
within each critical band of interest, in order to diffuse
or perturb the cancellation patterns and smooth the
perceived frequency response.

The alternating phase bands are spaced linearly, with a
band spacing of

$$b = \frac{f_s}{2N}, \qquad (8)$$

where f_s is the sample rate in Hz. The Equivalent
Rectangular Bandwidth (*ERB*) of the auditory filters is
approximated by

$$ERB = 24.7(0.00437F + 1), \qquad (9)$$

where F is the center frequency in Hz [11]. Assuming
the lowest critical band of interest is centered near the
lowest comb filter notch, which in our example is
around 2 kHz, the smallest *ERB* of interest would be
about 241 Hz. In order for the width b of our alternating
phase bands to be less than the *ERB*, we have

$$N > \frac{f_s}{2 \cdot 24.7(0.00437F + 1)}. \qquad (10)$$

In this case, given a 48 kHz sampling rate, the delay N
would be at least 100 samples, or about 2 ms.

3.6.1. Temporal Smearing

We would prefer to avoid unnecessarily long values of N that might cause perceptible temporal smearing of impulsive sounds, due to long decay times. If we view the impulse response as a type of reverberation, the reverberation time is given by

$$T_r = \frac{-60N}{g_{dB} \cdot f_s} \text{, where} \qquad (11)$$

T_r is the -60 dB reverberation time in seconds,
N is the length of the delay in samples,
f_s is the sample rate in Hz, and
g_{dB} is the allpass gain g expressed in dB.

Therefore, the reverberation time is proportional to the delay time and inversely proportional to the log of the gain.

With $N = 100$, and $g = 0.414$, the -60 dB reverberation time T_r is about 16 ms. This is a short decay time compared to that of most rooms, so the temporal smearing is unlikely to be perceptible over speakers with typical voice or music recordings.

The temporal smearing is, however, apparent with headphones, given impulsive inputs. The impulse response sounds slightly "fatter" or doubled. With music over headphones, a possible change in the timbre of snare drums was noted, along with a not unpleasant increase in spatial width.

If a reduction in temporal smearing is desired, it may prove useful to cancel all non-zero values of the impulse response after a specified delay by adding a scaled, delayed complementary allpass response to each of the original allpass filters [33]. However, the magnitude response will no longer be allpass.

3.7. Total Magnitude Response of Phase Diffusion Plus Acoustic Crosstalk

At this point, we would like to consider the ear signals resulting from the overall phase diffusion system, including the effect of the interaural crosstalk. Starting with a mono source, x, the signals B_L and B_R at the left and right ears are given by

$$\begin{bmatrix} B_L \\ B_R \end{bmatrix} = \begin{bmatrix} H_{LL} & H_{RL} \\ H_{LR} & H_{RR} \end{bmatrix} \begin{bmatrix} A_L \\ A_R \end{bmatrix} [x]. \qquad (12)$$

The total magnitude response at one ear, using our simple time delay model, is shown by the solid line in Figure 10. Note that the phase diffusion helps fill in the notches (dotted line) caused by the acoustic crosstalk. While the diffusion introduces new notches, these are smaller and closer together and will be smoothed by the auditory filters.
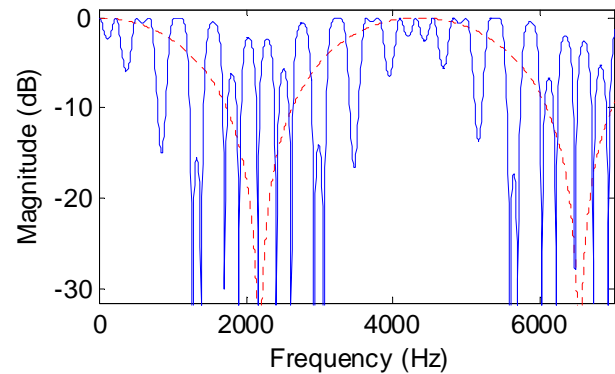

Figure 10. Magnitude response of simple delay models of the acoustic crosstalk at one ear, with speakers at ±30°: acoustic crosstalk alone (dotted); phase diffusion (N=100) plus acoustic crosstalk (solid).

Smaller values of the gain g, such as $g = 0.1$, do not provide much diffusion depth, so the resulting magnitude response tends to consist of small fluctuations around the original response. On the other hand, large values of g, such as $g = 0.8$, tend to produce unwanted dips where the original magnitude response had peaks. In addition, large values of g cause longer decay times and temporal smearing.

3.8. Spatial Smearing

Even though temporal smearing may not be apparent, there can be noticeable spatial smearing. With voice, for example, there can be a considerable increase in perceived image width, as might be expected considering the algorithm's origins as a pseudo-stereo circuit.

Since the ear's use of phase as a localization cue is primarily limited to frequencies below about 1 kHz, and since our goal is to diffuse the left and right phases

around the lowest comb filter notch frequency (2 kHz) and above, we can use a pair of complementary crossover filters, as shown in Figure 11, to limit our phase diffusion to frequencies above a selected crossover frequency, f_c.
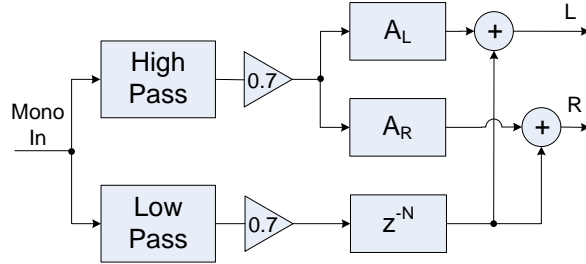

Figure 11. Complementary crossover filters limit the phase diffusion to the higher frequencies.

The $\sqrt{0.5}$ gains, shown as 0.7, are intended to normalize the reverberant sound pressure produced by the pair of allpass filters. The N-sample delay, z^{-N}, is used to match the average delay produced by allpass filters A_L and A_R.

The crossover helps reduce the increase in apparent image width, because we are no longer diffusing the phases in the low frequency range where phase is a primary localization cue. In practice, a slight spreading or pseudo-stereo effect may still be apparent, especially when the speakers subtend an angle of greater than ±60°, but the widening is subtle, and not unpleasant, for the smaller angles typically used for television viewing.

For listeners to the left or right of the line of symmetry between the speakers, this widening causes the image's pull toward the nearest speaker (as discussed in section 1.1) to be somewhat less obvious. While the phantom image is still not centered between the speakers, it is no longer so tightly focused toward one side.

3.9. Magnitude-Complementary Crossover Filters

When power-complementary crossover filters are used with the system of Figure 11, undesired fluctuations of the power response are noted in the vicinity of the crossover frequency, due to the interaction with the allpass filters. These can be minimized by using magnitude-complementary filters, which have matching phase responses at all frequencies.

A pair of digital filters having transfer functions $G(z)$ and $H(z)$ are defined to be magnitude-complementary if

$$|G(z)| + |H(z)| = 1, \tag{13}$$

or, equivalently, if the sum of their magnitude responses equals unity on the unit circle [34][35][36]; i.e.,

$$|G(e^{j\omega})| + |H(e^{j\omega})| = 1, \ 0 < \omega < 2\pi. \tag{14}$$

If we define two transfer functions $E(z)$ and $F(z)$ based on the sum and difference of two allpass functions [34],

$$E(z) = 0.5[A_1(z) + A_2(z)] \text{ and} \tag{15}$$

$$F(z) = 0.5[A_1(z) - A_2(z)], \tag{16}$$

where $A_1(z)$ and $A_2(z)$ are stable allpass transfer functions and $E(z)$ is a lowpass prototype filter, and define transfer functions $G(z)$ and $H(z)$, such that

$$G(z) = E^2(z), \text{ and} \tag{17}$$

$$H(z) = -F^2(z), \tag{18}$$

then [35],

$$\begin{aligned} G(z) + H(z) &= E^2(z) - F^2(z) \\ &= [E(z) + F(z)][E(z) - F(z)] \\ &= A_1(z)A_2(z). \end{aligned} \tag{19}$$

Therefore, $G(z)$ and $H(z)$ are allpass-complementary.

Since

$$\begin{aligned} &|G(e^{j\omega})| + |H(e^{j\omega})| \\ &= |E(e^{j\omega})|^2 + |F(e^{j\omega})|^2 = 1, \end{aligned} \tag{20}$$

the transfer functions $G(z)$ and $H(z)$ are also magnitude-complementary. Furthermore, it can be shown that phasors $G(e^{j\omega})$ and $H(e^{j\omega})$ have the same phase angles at all frequencies [35].

With magnitude-complementary filters, the outputs cross over at

$$\left|G\left(e^{j\omega_0}\right)\right| = \left|H\left(e^{j\omega_0}\right)\right| = 0.5, \qquad (21)$$

or -6 dB, instead of at -3 dB as with power-complementary filters [35].

To implement these filters, we note that the lowpass response is

$$G(z) = 0.5^2 \left[A_1(z) + A_2(z)\right]^2$$
$$= 0.25 \left[A_1^2(z) + 2A_1(z)A_2(z) + A_2^2(z)\right], \quad (22)$$

and the highpass response is

$$H(z) = -0.5^2 \left[A_1(z) - A_2(z)\right]^2$$
$$= -0.25 \left[A_1^2(z) - 2A_1(z)A_2(z) + A_2^2(z)\right]. \quad (23)$$

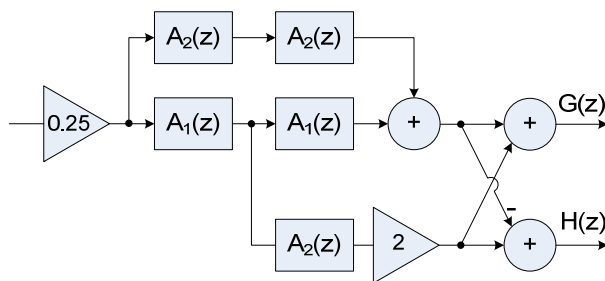An efficient implementation is shown in Figure 12.

Figure 12. Efficient implementation of a magnitude-complementary filter pair, after [37].

A simplified version of this topology was discovered in 1958 by Schroeder [13], who presented a linear-phase implementation in which $A_1(z)$ was a delay line and $A_2(z)$ was a pass-through. His goal was to create a circuit with no phase differences between the output signals, as part of his attempt to determine whether the pseudo-stereo effect was primarily a result of the complementary magnitude responses or the differential phase response [13]. (He later concluded that both magnitude and phase response differences could be used to produce pseudo-stereo effects [31].)

3.10. Approximate Allpass and Power Complementarity of the Outputs

Allpass complementarity, or an approximation thereof, is useful at low frequencies so the sum of the output signals will add in-phase to produce a (possibly delayed or allpassed) version of the input signal. Approximate power-complementarity is useful at mid- and high-frequencies, where the output signals are considered to have random phase and therefore add in terms of energy instead of amplitude.

Figure 13 illustrates the extent to which the outputs of Figure 11 are allpass- and power-complementary, given $N = 100$ and $f_c = 1.5$ kHz, with fifth-order magnitude-complementary lowpass and highpass filters $G(z)$ and $H(z)$, respectively, which in turn are comprised of second- and third-order allpass filters $A_1(z)$ and $A_2(z)$.
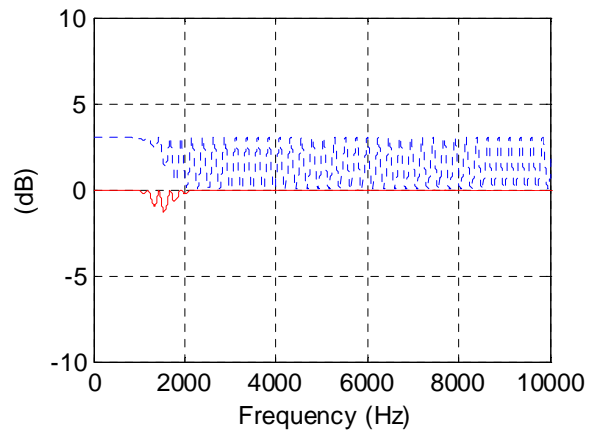

Figure 13. Allpass complementarity (dotted), and power complementarity (solid).

At high frequencies, the outputs are approximately power-complementary, as desired (shown by the solid line). There is about 1 dB of ripple in the power complementarity near the crossover frequency. This ripple would have been approximately twice as large if the crossover filters themselves had been power-complementary instead of magnitude-complementary. Higher-order crossover filters can be used to confine the ripple to a narrower range of frequencies.

At low frequencies, where signals are presumed to add in-phase, the outputs are approximately allpass-complementary (as shown by the dotted line), though with a 3 dB gain. This low-frequency boost could be eliminated by changing the gain following the lowpass

filter from $\sqrt{0.5}$ to 0.5; however, while the latter gain choice might be appropriate for use in an anechoic chamber, in a reverberant environment this would result in a 6 dB energy loss at low frequencies, as becomes apparent when comparing the timbre of pink noise signals. The low-frequency gain value of $\sqrt{0.5}$ preserves the perceived timbre in typical listening spaces, by ensuring that the overall system is more-or-less energy-preserving at all frequencies.

The 3 dB ripple in allpass complementarity above the crossover frequency is not relevant, since we assume that signals at these frequencies mix according to power instead of amplitude. In fact, this ripple is what helps break up the phantom mono phase cancellations, creating a flatter perceived response.

Over headphones, the $\sqrt{0.5}$ lowpass gain remains appropriate, since all that matters is that the left and right outputs are power-complementary. In this case, allpass complementarity is irrelevant, because the left and right signals do not get summed in the air.

3.11. Phase Response

In Figure 14, the two outside, DNA-like curves represent the phase responses of the left and right outputs of Figure 11, while the center curve is the phase response of the output of the lowpass filter plus N-sample delay. This delay is included in order to compensate for the average phase response of the allpass filters and prevent unnecessary cancellations. It is apparent from Figure 14 that the phase diffusion is limited to the higher frequencies. This helps disrupt the phantom mono phase cancellations without adversely affecting low frequency phase-based spatial cues.
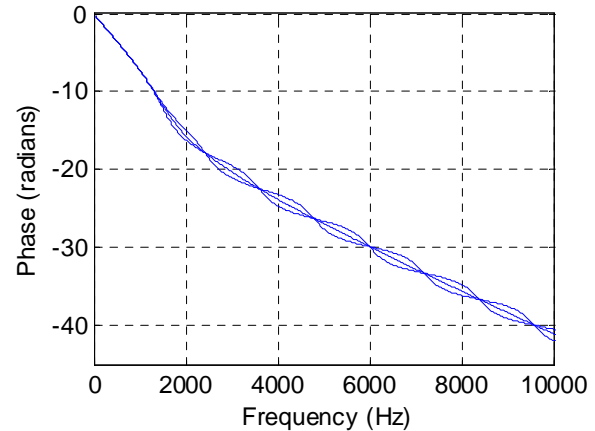

Figure 14. Phase responses, for $N = 20$: left and right outputs (outside curves), lowpass filter plus N-sample delay (inside curve).

The values of gain g, delay N, and crossover cutoff frequency f_c can be tuned as desired to balance the various perceptual effects.

3.12. Phase Diffusion for Stereo Inputs

The above systems are designed to convert a mono input to left and right outputs, in order to produce an improved phantom center. A simple way of adapting this system to work with stereo inputs is shown in Figure 15. Here, the left and right inputs are processed separately, with allpass filters applied to the highpass-filtered signals.

Note that allpass filters A_L and A_R are different, as in (6) and (7). As a result, any high-frequency content common to the left and right channels will be processed by A_L for one output and by A_R for the other, as in Figure 11, resulting in the interweaving phase responses shown in Figure 14.
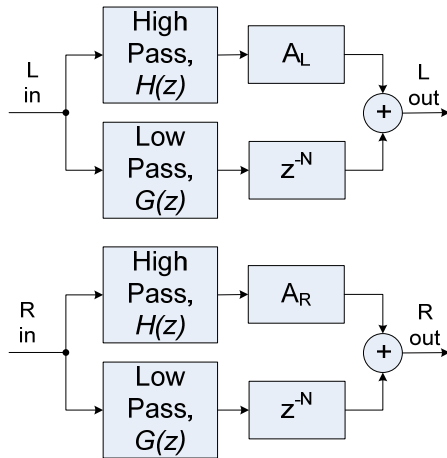
Figure 15. High-frequency phase diffusion
for stereo inputs.

A variety of phase-based decorrelation methods can be substituted for A_L and A_R. As before, the delay z^{-N} is used to match the average delay produced by allpass filters A_L and A_R.

3.13. Summary of the Method

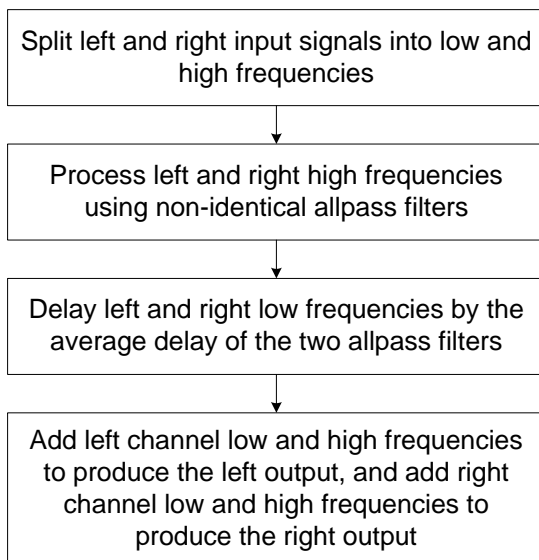The system shown in Figure 15 is summarized in the following flow chart.

Figure 16. Flow chart of the high-frequency
phase diffusion method for stereo inputs.

4.   RESULTS

The current method helps fill in the gaps caused by phase cancellations, within the resolution of the ear's auditory filter. As mentioned, in most typical consumer listening environments, problems due to notches in the magnitude response are rarely obvious except during controlled intelligibility tests or in very dry environments, because of the diffusion caused by room reflections. Consequently, the improvements also tend to be subtle.

During informal listening tests in an acoustically live office, no obvious improvement in dialogue clarity was noticed. However, this method did cause a reduction in the noticeability of coloration changes when moving the head from side to side. It also produced a slight improvement of the problem of phantom images moving toward the nearest speaker, by slightly blurring the apparent position so it was not so tightly focused. The allpass gain g controls the tradeoff between image focus and reduction of coloration.

The improvements were more noticeable in a relatively "dry" audio lab. Given a dual mono speech signal mixed with stereo "babble" noise, participants reported that when the effect was enabled, the intelligibility was slightly improved and the voice was somewhat clearer. In addition, the phantom center only started to collapse toward the nearest speaker when listeners moved to the extreme left or right; this may be advantageous in typical listening situations, since people rarely sit at exactly the same distance from the two speakers. Overall, the sound was described as broader, more stable and slightly further away.

With headphone listening, the center image was larger and not as oppressively focused at a single spot in the center of the head.

4.1.   Increase in Perceived Distance

As mentioned, a small increase in apparent distance was noticed, even when using values of delay N as short as 0.1 ms., or 5 samples at 48 kHz. At times, it seemed as if the diffusion could be perceived as either an increase in width or an increase in distance.

Corteel et al observed a similar increase in apparent distance, noticeable only at high frequencies; they hypothesized that diffusion reduces the precision of localization cues and that this reduction may be

interpreted as an increase in distance. They suggested that this effect called for further study [25].

Typically, the reverberant-to-direct energy ratio is regarded as a primary distance cue, with a larger percentage of reverberant energy implying a greater distance [11][25]. If we regard the first sample of the impulse response from the system of Figure 6 as the direct sound and all the following energy as the reverberation, the computed reverberant-to-direct energy ratio actually decreases with an increase in allpass gain g, as shown in the solid curve of Figure 17. However, preliminary listening tests suggest that for a constant delay N, the apparent distance increases as g goes from 0 to 0.7; thus, the apparent distance increases as the reverberant-to-direct ratio decreases. This is the opposite of what we would expect.
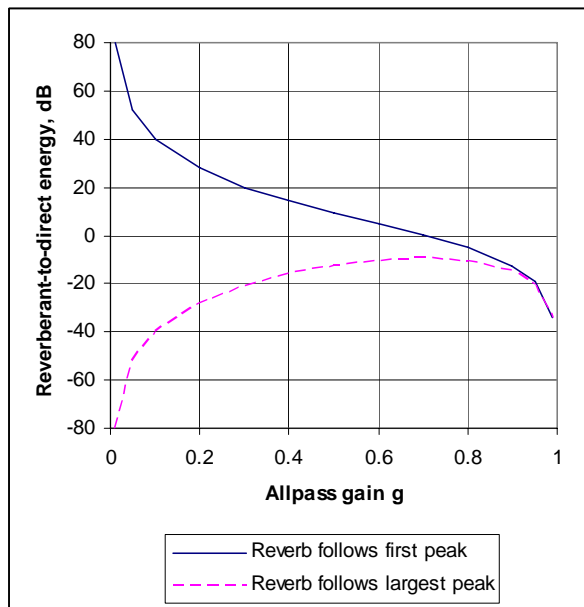

Figure 17. Reverberant-to-direct ratios, in dB, for various magnitudes of gain g, regarding the reverberant response as everything following the first peak (solid), or everything following the largest peak (dashed).

Note, however, that the second peak of the impulse response of an allpass filter is often greater in magnitude than the first. It seems likely that the peak with the largest magnitude may be perceived as being part of the direct path. If we assume that, perceptually, the reverberation starts after the largest peak, we get the reverberant-to-direct ratios shown in the dashed curve of Figure 17.

In this case, the reverberant-to-direct energy ratio reaches a maximum at an allpass gain of $g = \sqrt{0.5}$. Informal listening tests suggest that the perceived distance reaches a maximum when g is approximately 0.7 and decreases above and below that value. This agrees with the idea that greater reverberant-to-direct energy ratios correspond to greater perceived distances, so long as we define the reverberant energy as starting after the largest peak of the impulse response.

In general, the increase in perceived distance is slight and unobjectionable. Furthermore, a similar increase in perceived distance, together with similar temporal and spatial smearing, is expected to result from simply playing the original dual mono signal in a reverberant room. If we want to achieve the desired effects while limiting any increase in perceived distance, it might be worthwhile to explore ways of maximizing the inter-channel phase differences while minimizing the reverberant-to-direct energy ratio of the decorrelation method.

4.2. Advantages over Other Methods

Systems that decorrelate audio by creating magnitude differences in alternating frequency bands (for example, using pseudo-stereo comb filters) would create timbre problems for listeners located closer to one speaker.

Decorrelating the left and right signals simply by adding early reflections or reverberation might unnecessarily color the frequency response or lengthen the impulse response. In addition, without the crossover filters of Figure 11, the resulting full-spectrum decorrelation would impose unwanted phase changes at low frequencies, where phase information is important for localization.

Finally, without using in-phase magnitude-complementary crossover filters, there would be significant ripples in the power response near the crossover frequency.

5. DISCUSSION

We reviewed a number of problems relating to the phantom center image and presented a system which improves many of these problems by decorrelating the left- and right-channel phases at higher frequencies.

The current method may help reduce the perception of comb filter coloration changes that occur when moving the head. In addition, it is possible that it may slightly enhance dialogue intelligibility, especially in acoustically dry environments. The mild spatial blurring helps make the collapse of the phantom image toward the nearest speaker somewhat less obvious, and it greatly improves the problem of headphone compatibility by spreading the center image.

Above the crossover frequency, the current method also eliminates the 3 dB increase in center balance and the 3 dB decrease in reverberant-to-direct ratio associated with mixing to mono.

5.1. Possible Applications

The current phase decorrelation method may prove most useful for asymmetrical listening geometries and in relatively non-reflective acoustical environments such as outdoor amphitheatres, as in the Hollywood Bowl experiments mentioned in section 3.1. Since the phantom center image in acoustically dry environments suffers from spectral coloration and related problems, this method can be seen as compensating for the lack of room reflections without interfering with low-frequency localization

This method may also have interesting applications for multiple-speaker systems such as wave field synthesis, ambisonics, beam-forming, 5.1-speaker systems, etc. In the context of wave field synthesis, Corteel et al found that diffuse filtering may reduce spatial aliasing artifacts [25]. The current research may also help minimize the coloration or phasiness of multiple-speaker "soundbar" systems used for television.

5.2. Additional Research

Formal listening tests are needed to determine the extent of the improvement, if any, to the intelligibility of speech (particularly in the presence of uncorrelated noise). Tests in a variety of environments, with different types of audio content, should be used to investigate the other improvements mentioned and highlight any unwanted artifacts. In particular, this method should be carefully evaluated for any impairments of the stereo image at higher frequencies and any perceived separation of the low- and high-frequency components of individual musical instruments.

An interesting experiment would be to perform an intelligibility test similar to that of Shirley et al [9], using phantom center vs. real center speakers in an anechoic chamber. To the extent that the comb filter notches are responsible for the reduced intelligibility of the phantom center case, we would expect the results to be worse in the anechoic chamber, since the notches would not be filled in by room reflections. Listening with only one ear would eliminate the averaging of the spectra at the two ears, further highlighting the effect. Pink noise or babble noise from a real center speaker, not subject to the phantom center notches, could be used to impair the signal-to-noise ratio in the vicinity of the notches, improving the sensitivity of the test. (Alternatively, decorrelated stereo noise could be used, but it might be less effective due to spatial unmasking.) Afterwards, the results from standard dual mono could be compared to those obtained from the current phase decorrelation algorithm.

Other research could involve comparing various decorrelation methods and parameter values in order to maximize the desired results (reduced coloration, elimination of the Haas effect, and enhanced dialogue clarity) while minimizing undesired effects such as the increase in perceived distance.

Finally, it might be interesting to compare estimates of perceived localization and coloration for various input and output signals using a computational model of binaural auditory perception, such as that of Pulkki et al [38].

6. ACKNOWLEDGMENT

7. REFERENCES

[1] Blumlein, A.D., "Improvements in and relating to sound-transmission, sound-recording and sound-reproducing systems," British Patent 394,325. Reprinted in *J. Audio Eng. Soc.*, vol. 6, pp. 91-98, 1958.

[2] Jens Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization*, MIT Press, Cambridge, MA, 1997 (original edition, 1974).

[3] Benjamin Bauer, "Phasor Analysis of Some Stereophonic Phenomena," *IRE Transactions on Audio*, vol. 10, Issue 1, Part 1, pp. 18-21, 1962, January.

[4] Mark Gardner, "Some Single- and Multiple-Source Localization Effects," *J. Audio Eng. Soc.*, vol. 21, no 6, pp 430-437, 1973, July/August.

[5] J.M. Eargle, "Stereo / Mono Disc Compatibility: A Survey of the Problems," *J. Audio Eng. Soc.*, vol. 17, no. 3, pp. 276-281, 1969 June.

[6] Tomlinson Holman, "New Factors in Sound for Cinema and Television," *J. Audio Eng. Soc.*, vol. 39, no. 7/8, pp. 529-539, 1991 July/August.

[7] Benjamin Bauer, "Some Techniques toward Better Stereophonic Perspective," *IEEE Trans. on Audio*, AU-11, 88, 1963 May-June. Reprinted in *J. Audio Eng. Soc.*, vol. 17, no. 4, pp. 410-415, 1969 August.

[8] Benjamin Bauer, "Improving Headphone Listening Comfort," *J. Audio Eng. Soc.*, vol. 13, no. 4, pp 300 – 302, 1965, October.

[9] Ben Shirley, Paul Kendrick, and Claire Churchill, "The Effect of Stereo Crosstalk on Intelligibility: Comparison of a Phantom Stereo Image and a Central Loudspeaker Source," *J. Audio Eng. Soc.*, vol. 55, no. 10, pp. 852-863, 2007 October.

[10] Richard O. Duda, William L. Martens, "Range Dependence of the Response of a Spherical Head Model," *J. Acoust. Soc. Am.*, vol. 104, no. 5, pp. 3048 – 3058., 1998 November.

[11] Brian Moore, *An Introduction to the Psychology of Hearing*, London, UK, Academic Press, 1997.

[12] David Clark, "Measuring Audible Effects of Time Delays in Listening Rooms," presented at the AES 74th Convention, New York, NY, paper 2012, 1983 October 8-12.

[13] M.R. Schroeder, "An Artificial Stereophonic Effect Obtained from a Single Audio Signal," *J. Audio Eng. Soc.*, vol. 6, no. 2, pp. 74-79, 1958 April. Republished in (ed.) J. Eargle, *Stereophonic Techniques*, Audio Engineering Society, New York, 1986.

[14] Francis Rumsey, *Spatial Audio*, Oxford, UK, Focal Press, 2001.

[15] R. Bücklein, "The Audibility of Frequency Response Irregularities," *J. Audio Eng. Soc.*, vol. 29, no. 3, pp. 126-131, 1981 March.

[16] Anton Ehrenzweig, *The Psycho-Analysis of Artistic Vision and Hearing: An Introduction to a Theory of Unconscious Perception*. New York: Julian Press, 1953.

[17] Günther Theile, "On the Localisation in the Superimposed Soundfield," dissertation submitted for the degree Dr.-Ing. Technische Universität Berlin (translated by Tobias Neher), 1980.

[18] Helmut Wittek, Francis Rumsey, and Günther Theile, "On the Sound Colour Properties of Wavefield Synthesis and Stereo," presented at the AES 123rd Convention, New York, NY, paper 7167, 2007 October 5-8.

[19] William Gardner, "Reverberation Algorithms," in *Applications of Digital Signal Processing to Audio and Acoustics*, edited by Mark Kahrs and Karlheinz Brandenburg, Norwell, MA: Kluwer Academic Publishers, 1998.

[20] Dave Moulton, "The Center Channel: Unique and Difficult," http://www.tvtechnology.com/pages/s.0065/t.1349.html.

[21] M.R. Schroeder, and B. S. Atal, "Computer Simulation of Sound Transmission in Rooms," *IEEE International Convention Record* (7), New York: IEEE Press, 1963.

[22] Jerry Bauck, "Equalization for Central Phantom Images and Dependence on Loudspeaker Spacing: Reformatting from Three Loudspeakers to Two Loudspeakers," presented at the AES 109th Convention, Los Angeles, CA, paper 5240, 2000 September 22-25.

[23] Jean-Marc Jot, *Etude et Réalization d'un Spatialisatuer de Sons par Modèles Physiques et Perceptifs*, PhD thesis, France Telecom, Paris, 92 E 019, 1992 (in French).

[24] Jean-Marc Jot and Antoine Chaigne, "Digital Delay Networks for Designing Artificial Reverberators," presented at the AES 90th Convention, paper 3030, 1991 February.

[25] Etienne Corteel, Khoa-Van NGuyen, Olivier Warusfel, Terence Caulkins, Renato Pellegrini, "Objective and Subjective Comparison of Electrodynamic and MAP Loudspeakers for Wave Field Synthesis," presented at the AES 30th International Conference, Saariselkä, Finland, 2007 March 15-17.

[26] Matti Karjalainen and Hanna Järveläinen, "Reverberation Modeling using Velvet Noise," presented at the AES 30th International Conference, Saariselkä, Finland, 2007 March 15-17.

[27] Yi-Wen Liu and Julius Smith III, "Perceptually Similar Orthogonal Sounds and Applications to Multichannel Acoustic Echo Canceling," presented at the AES 22nd International Conference on Virtual, Synthetic and Entertainment Audio, Espoo, Finland, 2002 June 15-17.

[28] Gary Kendall, "The Decorrelation of Audio Signals and Its Impact on Spatial Imagery," *Computer Music Journal*, 19:4, pp. 71-87, 1995 Winter.

[29] Maurice Bouéri and Chris Kyriakakis, "Audio Signal Decorrelation based on a Critical Band Approach," presented at the AES 117th Convention, San Francisco, CA, USA, 2004 October 28-31.

[30] George Augspurger, Søren Bech, Rollins Brook, Elizabeth Cohen, Thomas Schindler, and John Eargle, "Use of Stereo Synthesis to Reduce Subjective/Objective Interference Effects: The Perception of Comb Filtering, Part II," presented at the AES 87th Convention, New York, NY, 1989 October 18-21.

[31] M. R. Schroeder, "Improved Quasi-Stereophony and 'Colorless' Artificial Reverberation," *J. Acoustical Soc. Of America*, vol. 33, no. 8, 1961 August.

[32] Henrik Møller, Pauli Minnaar, Søren Olesen, Flemming Christensen, and Jan Plogsties, "On the Audibility of All-Pass Phase in Electroacoustical Transfer Functions," *J. Audio Eng. Soc.*, vol. 55, no. 3, pp. 115-134, 2007 March.

[33] Renato Pellegrini, "Perception-based Design of Virtual Rooms for Sound Reproduction," AES 22nd International Conference on Virtual, Synthetic and Entertainment Audio, 2002 June.

[34] S. K. Mitra, *Digital Signal Processing: A Computer-Based Approach*. New York: McGraw-Hill, 2006.

[35] Phillip A. Regalia and Sanjit K. Mitra, "A Class of Magnitude Complementary Loudspeaker Crossovers," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-35, no. 11, pp. 1509-1516, 1987 November.

[36] S. Radhakrishnan Pillai and Gregory H. Allen, "Generalized Magnitude and Power Complementary Filters," *Proc. IEEE Int. Conf. ASSP*, vol. 3, pp. 585-588, 1994, 19-22 April.

[37] Pedro Reviriego Vasallo, José Parera, and Ramón García Gómez, "Linear-Phase Crossover Design using Digital IIR Filters," *J. Audio Eng. Soc.*, vol. 46, no. 5, 1998 May.

[38] Ville Pulkki, Matti Karjalainen, and Jyri Huopaniemi, "Analyzing Virtual Sound Source Attributes using a Binaural Auditory Model," *J. Audio Eng. Soc.*, vol. 47, no. 4, pp. 203-217, 1999 April.

[39] Anderson, Joseph L., "Classic Stereo Imaging Transforms — A Review," http://www.dxarts.washington.edu/courses/567/08WIN/JL_Anderson_Stereo.pdf.