SIGNAL DECORRELATION USING PERCEPTUALLY INFORMED ALLPASS FILTERS

Elliot Kermit-Canfield and Jonathan Abel

Center for Computer Research in Music and Acoustics,
Stanford University, Stanford, CA 94305 USA
`[kermit|abel]@ccrma.stanford.edu`

ABSTRACT

When a monophonic source signal is projected from two or more loudspeakers, listeners typically perceive a single, phantom source, positioned according to the relative signal amplitudes and speaker locations. While this property is the basis of modern panning algorithms, it is often desirable to control the perceived spatial extent of the phantom source, or to project multiple, separately perceived copies of the signal. So that the human auditory system does not process the loudspeaker outputs as a single coherent source, these effects are commonly achieved by generating a set of mutually decorrelated (e.g., statistically independent) versions of the source signal, which are then panned to make an extended source or multiple, independent source copies.

In this paper, we introduce an approach to decorrelation using randomly generated allpass filters, and introduce numerical methods for evaluating the perceptual effectiveness of decorrelation algorithms. By using allpass filters, the signal magnitude is preserved, and the decorrelated copies and original signal will be perceptually very similar. By randomly selecting the magnitude and frequency of the poles of each allpass biquad section in the decorrelating filter, multiple decorrelating filters may be generated that maintain a degree of statistical independence. We present results comparing our approach (including methods for choosing the number of biquad sections and designing the statistics of the pole locations) to several established decorrelation methods discussed in the literature.

1. INTRODUCTION

Signal decorrelation is an important tool for audio upmixing, spatialization, and auraliztion. In the simplest case, when two coherent audio signals are played through loudspeakers, a listener will perceive a single sound source located somewhere between the two speakers, controlled by the relative amplitudes and time delay of the signals. Although this is one of the principles on which stereophonic panning algorithms rely, it is not without problems. For example, when the signals are presented to the listener over headphones, the location of the source is often perceived to be within the listener's head. Additionally, the perceived source width is often reduced. These problems also exist in systems with more than two loudspeakers.

Our goal in decorrelating signals is to reduce the phase coherence of a given signal while maintaining perceptual transparency. That is to say a monophonic file run though our decorrelation algorithm presented on a single speaker should be indistinguishable from the original file. When the same file is run through multiple independent decorrelation filters and presented on multiple speakers, the signals should no longer sum perceptually to a single point and the apparent source width should appear to be extended. See Fig. 1 for a graphical depiction of a phantom source

between two speakers and decorrelated phantom sources. In the current work, we are concerned with applications for multichannel and surround sound applications where one might want many decorrelated copies of a signal panned in space.

In the following sections we will briefly review other common decorrelation techniques followed by a description of the measures we use to evaluate perceptual transparency and independence. We will then introduce our approach to signal decorrelation through perceptually-weighted random allpass filters before showing the results of evaluating our approach in comparison to other popular techniques. We will conclude with some recommendations for using decorrelation filters in real-time systems and the implications of this work for further research.
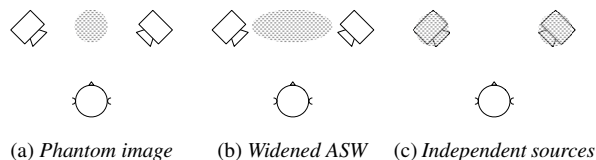

(a) *Phantom image* (b) *Widened ASW* (c) *Independent sources*

Figure 1: *Multiple speakers producing the same signal usually produce a single phantom image (1a) while decorrelation can produce a phantom image with a wider apparent source width (1b) or multiple, statistically independent sources (1c).*

2. STANDARD DECORRELATION TECHNIQUES

Perhaps the simplest decorrelation algorithm is based on convolving the source signal with a short sequence of random samples, scaled to have unit power. The longer we make this sequence the more decorrelation we can achieve, but at the expense of smearing the original signal out in time. For this reason, we typically constrain the length to be shorter than $30\,\mathrm{ms}$ so the decorrelation is not perceived to add reverberation.

In a second approach Kendall proposes an "allpass filter" technique formed by taking the IDFT of a transform constrained to have unit magnitude at the DFT bins with a random phase uniformly distributed on the interval $(-\pi, \pi)$ [1, 2]. Even though this processes creates a filter with unit magnitude *on* the DFT bins, it does not guarantee a flat magnitude response. Additionally, because the DFT is a periodic transform, discontinuities at the start and end of the signal can cause audible artifacts. It is possible to refine this technique by limiting how quickly the phase can change between consecutive samples, but this does not solve the issues related to the magnitude response being only flat on average. Due to the random nature of these algorithms, both techniques work when trying to decorrelate a signal into more than two channels.

Another broad approach involves passing the signal through a filterbank and applying delays to each band of frequencies [3–5]. Researchers have experimented with the number of frequency bands—ranging from as few as three to greater than 24—the spacing of the bands of the filterbank, and the amount of delay in each band. Depending on the complexity of the filterbank, this technique can suffer from frequency cancellation at the band edges during reconstruction, and can be computationally expensive. Furthermore, depending on the algorithm, some these techniques are not necessarily effective for multichannel applications as they take advantages of stereo, complementary processing.

Cabrera proposes a method of sinusoidal modeling with frequency and amplitude modulation to decorrelate signals [6,7]. This method introduces latency due to the signal analysis and can be computationally expensive.[1]

At the additional cost of higher complexity, many of these techniques can be further refined by separating the signal-to-be-processed into steady state and transient components, so that the decorrelation effects can be applied only to the steady state portion. This is done to ensure that transients do not get smeared out in time and become perceptual artifacts.

In addition to the aforementioned techniques, Gardner and Schroeder suggested methods of expanding the spatial extent of a signal using delays and comb filters, but these methods impart strong coloration on the signal [8,9].

The literature is diverse on decorrelation for upmixing and resynthesizing ambiance with applications for perceptual audio coders. Using a down-mixed, monophonic signal and side-chain information (binaural cue coding), Faller and Baumgarte suggest a perceptually-weighted frequency-domain modification using random sequences for each channel to preserve time and level differences [10, 11]. The MPEG Surround standard includes specification for an allpass filter approach [12, 13].

Valin proposes decorrelating frequencies above 2 kHz using shaped comb-allpass filters and lower frequencies by injecting psychoacoustically masked noise [14]. Zotter et al. take an approach using deterministic allpass filters [15].

Several mono-stereo upmixing techniques rely on leveraging complementary filters or other intrinsic properties of decorrelating only two channels [16–20]. While it is possible that these techniques can be extended for multichannel systems by cascading the algorithms with different parameters, this will not work for all systems (e.g., those require placing signals 90 degrees out of phase).

3. ALLPASS FILTERS FOR DECORRELATION

Allpass filters are useful for signal decorrelation because they maintain a flat frequency response while effecting the phrase and group delay. Digital allpass filters contain poles inside the unit circle matched with zeros at reciprocal magnitudes and at the same frequencies as the pole. In other words, the position of the zeros are reflected across the unit circle. The class of filters we consider are biquad allpass filters that produce a real output signal and have the

---

[1]Currently this does not run in real-time, but rather performs the analysis offline.

z transform

$$H(z) = \frac{\rho + z^{-1}}{1 - \overline{\rho}z^{-1}} \cdot \frac{\overline{\rho} + z^{-1}}{1 + \rho z^{-1}} \qquad (1)$$

$$= \frac{z^{-2} - 2\Re(\rho)z^{-1} + |\rho|^2}{1 - 2\Re(\rho)z^{-1} + |\rho|^2 z^{-2}}, \qquad (2)$$

where

$$\rho = \alpha e^{2\pi\omega j}, \qquad (3)$$

in which α controls the distance of the poles from the unit circle and ω is the angle. This can also be written as the difference equation

$$y[k] - 2\Re(\rho)y[k-1] + |\rho|^2 y[k-2] \qquad (4)$$
$$= x[k-2] - 2\Re(\rho)x[k-1] + |\rho|^2 x[k],$$

where $y[k]$ is the output and $x[k]$ is the input at sample k. These allpass filters exhibit conjugate symmetry, which causes their output to be real. This is important for processing real signals as it maintains the magnitude at all real frequencies. Here we propose cascading multiple, randomly generated allpass sections to generate decorrelating filters. Because they are randomly generated, we can create multiple, mutually decorrelated impulse responses with different phase responses, and therefore multiple decorrelated copies of signals.

In the following sections, we will explore the parameterization of this filter structure.

3.1. Pole-Zero Angle

The angle of the pole position controls the frequency at which the slope of the phase response changes fastest. When constructing our cascade of filters, one easy technique for choosing the pole's angle would be to place them randomly. While this is a valid technique, we have chosen to warp the random pole angles by equivalent rectangular bands (ERB) in order to maintain a relatively constant pole density across the critical bands of human hearing. As it turns out, decorrelation is difficult at low frequencies due to long wavelengths. This closely resembles the difficulty humans have localizing low frequency sound sources.

At high frequencies, the human auditory system primarily uses level differences to localize sound sources and decorrelating signals with time delays is potentially wasted [21]. High frequency signals have short wavelengths. Time delays could potentially realign signals' peaks and nulls offset by one or more periods. By warping the random distribution by ERBs, the majority of the poles will be placed in frequency ranges that will have the largest perceptual effect for sound localization.

3.2. Pole-Zero Radii

While the pole angle controls the frequency at which the largest phase change occurs, the pole's radius—the distance between the pole and the unit circle—controls the amount of the phase distortion. Each pole added to the system adds a cumulative π amount of phase to the system. As the pole's radius approaches the center of the unit circle, the phase across all frequencies approaches a linear slope and the response is identical to the phase response of a unit delay. As the pole radius approaches the unit circle, the phase distortion becomes concentrated in a smaller frequency region. Additionally, as the pole approaches the unit circle, the group
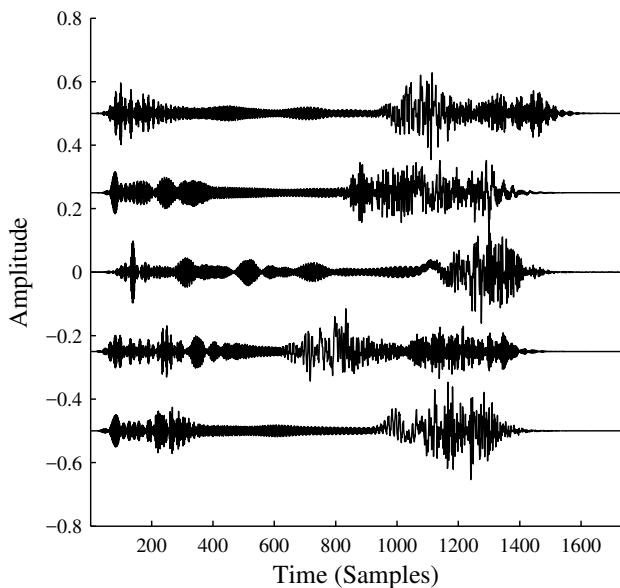
Figure 2: *Five allpass cascade impulse responses (1024 biquad sections), offset vertically.*

delay also becomes larger at that frequency. The pole radius will always be between $0 \leq r < 1$, and in the current approach, we select radii randomly within the limits

$$0.5 \leq r < \beta \,, \qquad (5)$$

where r is the radius and β is computed from

$$\beta = \frac{\tau_g}{f_s} \,, \qquad (6)$$

in which τ_g is the maximum allowable group delay and f_s is the sampling rate. We computed β to keep the group delay associated with any given pole below a perceptual threshold of $30\,\mathrm{ms}$ to minimize audible artifacts.

3.3. Filter Delay

When cascading the allpass filters, we can replace the unit delay, z^{-1}, with a longer delay, z^{-n}, in order to cause the phase to wrap around the unit circle n times faster. We randomly choose one, or small, prime or near prime delays so the cumulative effect of multiple delays do not align precisely.

Fig. 2 shows two example impulse responses generated by the above allpass filter cascade approach.

4. EVALUATION AND RESULTS

In order to evaluate the success of a decorrelation algorithm, we must consider the degree to which it achieves decorrelation, the amount it perceptually alters the original signal, and its computational efficiency. Correlation can be investigated mathematically through various applications of the cross-correlation function and by studying coherence. Perceptual similarity and the effects on localization are studied with empirical data from informal listening tests.

4.1. Cross-Correlation Metrics

The primary metric for evaluating the correlation between two discrete time signals is the cross-correlation function defined

$$\Phi_{xy}[l] = \sum_{m=0}^{N-1} x[m]y[l+m] \,, \qquad (7)$$

where x and y have been zero padded to be the same length and N is the number of samples in one of the signals. This function, measures the similarity between two signals as they are slid by each other at increasing time lags, l. When $x = y$, this function is called the autocorrelation function and will exhibit a maximum peak at lag zero ($l = 0$) with a height equivalent to the signal power. In the cross-correlation case, the more similar signals x and y are, the larger a peak will be seen in the cross-correlation. Additionally, when signals are similar but offset in time from one another, the maximum peak will shift away from lag zero by the amount of delay between the signals. It is important to note that correlation is only valid if the signals have similar features and are worth comparing. Finding the correlation of two unrelated signals will likely show that the signals are decorrelated but the result is not meaningful.

When we apply our decorrelation kernels to a signal, we hope to spread the energy of the signal out in time by different amounts across frequencies. Since our decorrelation algorithm is a LTI system, we can directly compare impulse responses generated by our allpass filter cascade. Fig. 3 shows an example of the autocorrelation and cross-correlation of two allpass decorrelation kernels. As we would expect, the autocorrelation of both impulse responses are highly correlated at lag zero and are not well correlated at any other delay. The cross-correlation of the two impulse responses, on the other hand, have no single point where they line up, and have no large spikes in their correlation.

Because the decorrelation is frequency-dependent, and most listeners have two spatially separated ears, it is imperative that we consider correlations at lags other than zero. Moreover, as we slide signals by each other in the computation of the cross-correlation, it is only at lag zero where signals will overlap entirely and all samples will be contributing to the cross-correlation function. To address these issues, we propose a cross-correlogram, computed by chunking the input signals into windows and performing 50% overlap-add cross-correlations on the windowed signals. This metric allows us to easily display cross and autocorrelations in a manner similar to the interaural cross-correlation (IACC).

In order to compare auto and cross-correlations, we normalize the range of the autocorrelation by the signal power and the cross-correlation by

$$\frac{1}{2} \Big(\max(|\Phi_{xx}|) + \max|\Phi_{yy}|) \Big) \,. \qquad (8)$$

This effectively scales the autocorrelation for both signals and the cross-correlation of the signals to each be within the range $[-1 \leq \Phi \leq 1]$, where -1 is perfect negative correlation and 1 is perfect correlation.

For visual clarity, let us consider a linear, sinusoidal chirp from $20\,\mathrm{Hz}$–$20\,\mathrm{kHz}$, seen in Fig. 4. Due to the periodicity of the signal increasing over time, there is a clear pattern imprinted in the auto-correlogram. Fig. 6 shows the auto and cross-correlogram of the present decorrelation technique as well as the noise and Kendall approaches. The distortion of the correlation patterns from the
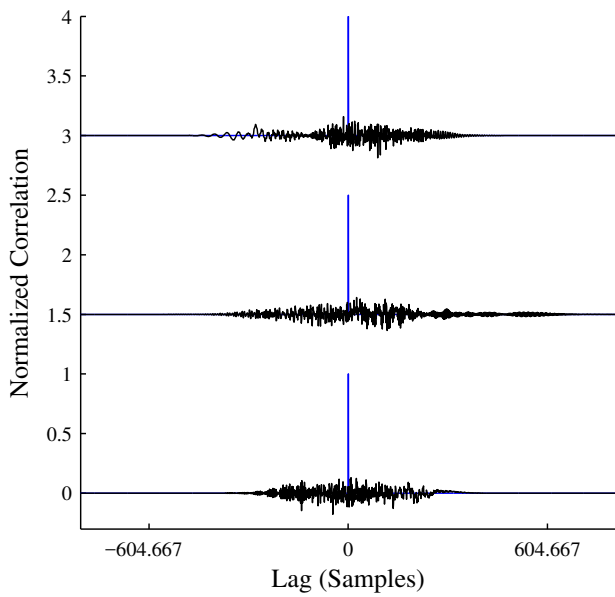
Figure 3: *Cross-correlation (black) and autocorrelation (blue) of three pairs of allpass cascade impulse responses (each 1024 biquads), offset vertically. Note the autocorrelation is a Dirac pulse.*

auto-correlogram are clearly seen in the cross-correlogram plots. It is important to note that the present technique preserves original input signals better than the other techniques, and this can be seen in the auto-correlogram by the fact that it is much more similar to the unfiltered auto-correlogram seen in Fig. 5.

4.2. Coherence

Coherence is defined as the maximum value of the cross-correlation function. Although coherence distills the correlation sequence to a single number, it is not directly useful as it treats all frequencies the same and does not take the human auditory system into consideration. Instead, we present coherence values for octave bands, computed using a 4th order zero-phase Butterworth filterbank. Ideally, the average coherence values across frequencies would be near zero. Fig. 7 shows coherence per octave band for the allpass filter cascade approach as well as the noise and Kendall methods. Because of the perceptually informed filter design, our technique achieves low coherence above 60 Hz but does not perform well near DC.

4.3. Efficiency

Allpass filters are inherently IIR filters because of the feedback associated with having poles. The random allpass filter cascade is precomputed and the filter is applied at runtime. While the class of decorrelation filter described in this paper could simply be implemented using their difference equations, we did most of our computation by approximating the filter's impulse response by thresholding it once it dropped below -90dB. Below this threshold, we assume the filter's output would be below the noise floor and indistinguishable from roundoff error. We applied the filter impulse response as a convolution kernel (implemented with multiplication in the frequency domain). These impulse responses are quite
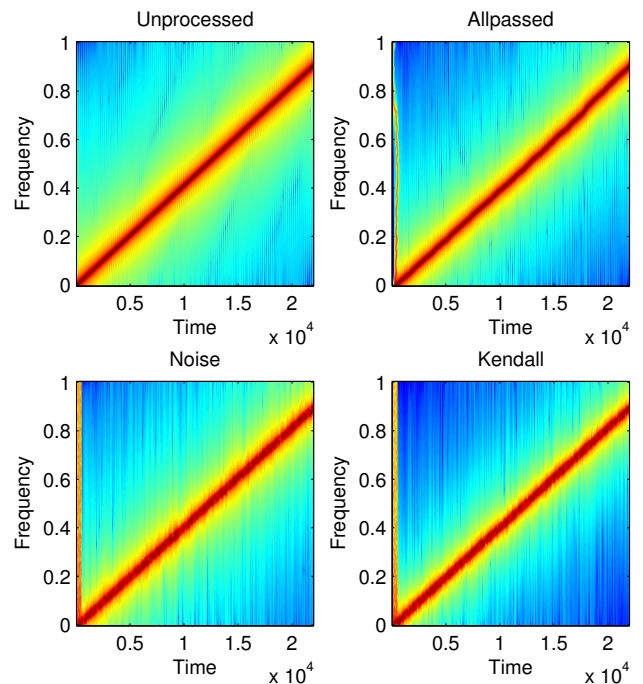

Figure 4: *Spectrogram of* 20 Hz–20 kHz *linear sinusoidal chirp depicting the unprocessed signal (a), allpass cascade (b), noise sequence (c) and Kendall sequence (d).*

short, having at most a few thousand taps and are suitable for real-time applications. For most applications, the computational requirements of this decorrelation algorithm will be negligibly small compared to other necessary signal processing constraints.

4.4. Listening Results

We performed informal listening tests to compare the allpass cascade decorrelation technique to the Kendall and noise approaches.[2] The authors and several colleagues familiar with audio and with normal hearing participated in a short listening test to evaluate the amount of decorrelation and the perceptual transparency of the decorrelation techniques. We presented critical listening material including castanets, glockenspiel, and male German speech from the EBU Sound Quality Assessment Material (SQAM) CD as well as recordings of unprocessed electric guitar and vocals from Tom's Dinner [22]. Each recording was presented over headphones and stereo studio monitors with listeners both in the sweet spot and off axis in a quiet room. The various conditions for the test included stereo presentations of the sound examples converted to mono and processed with 250, 500, 1000, 2000, and 3000 allpass biquad sections. As a control, we also processed the material using the Kendall and noise techniques with 512, 1024, and 2048 length sequences.

All three decorrelation algorithms achieved satisfactory decorrelation of the test stimuli. The noise and Kendall approaches decorrelated the signals slightly better than the allpass cascade, but the apparent source width of the signals for all three was significantly larger than the unprocessed "big" mono presentation. Like

---

[2]We intend to run formal listening tests in the future.
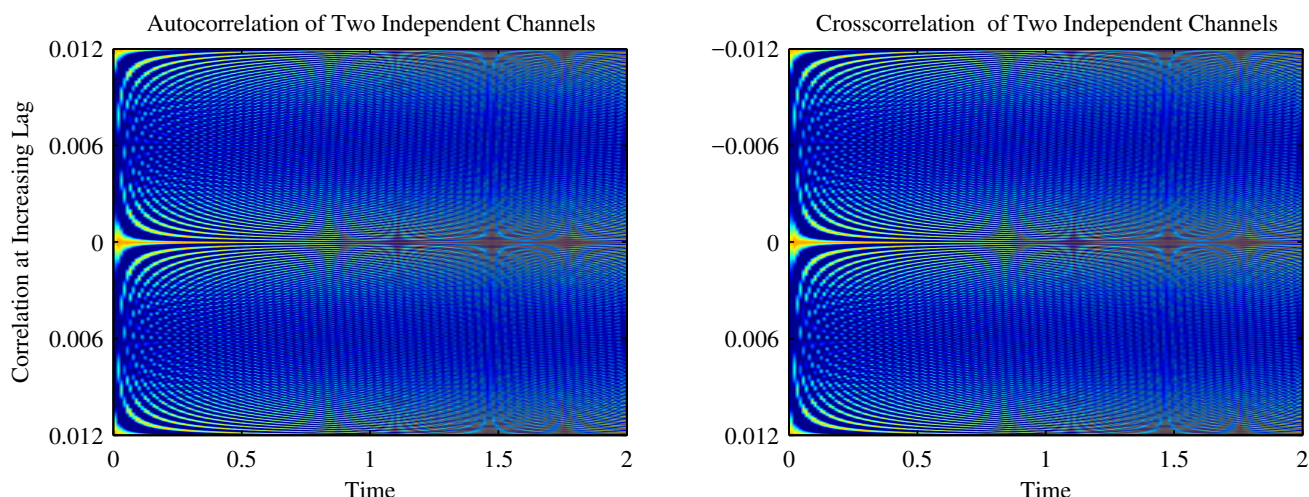
Figure 5: *Autocorrelelogram and cross-correlelogram for unprocessed* 20 Hz–20 kHz *linear sinusoidal chirp signal.*

when multiple violins play the same part in an orchestra and the similar sonic components fuse together to create the impression of a source with a larger apparent source width, the decorrelation algorithms broadened the spatial extent of the monophonic audio examples.

While all three algorithms effectively decorrelated the test material, the allpass cascade sounded much less colored than the other approaches. This is primarily due to the fact that the allpass filter maintain the desired flat frequency response while the Kendall and noise approaches do not. Presented in mono, the allpass cascade approach sounds much more similar to the original audio files and is therefore significantly more perceptually transparent. Presented in stereo, the Kendall and noise methods sound more similar to playing an audio signal combined with an inverted phase version of itself and the sonic coloration can be audible and unpleasant compared to the allpass filter cascade.

The optimal number of biquad sections for the allpass filter technique varies with the source material. In general, using more biquad sections reduces the correlation between channels. However, strong transients such as the onsets of the castanets and glockenspiel can suffer from chirp-like artifacts. In our tests, we found that the castanets were effected when more than 500 biquads were used and the glockenspiel at 1000. All the other material sufficiently masked the displeasing artifacts beyond 2000 biquad sections. Additionally, since the decorrelation filters are randomly generated, their effectiveness varies based on the frequency content of the material and the specific decorrelation filters used.

5. CONCLUSIONS

In this paper, we have described an approach for signal decorrelation that uses cascades of allpass filter biquad sections. Allpass filters are useful for decorrelation because they maintain a flat magnitude frequency response while adding frequency specific phase delays. We have introduced several methods for displaying and evaluating decorrelation, mainly the cross-correlogram, octave band coherence measure, and listening tests. While techniques like Kendall's "allpass filter" and convolution with noise can achieve a higher amount statistical independence, the present algorithm in-

troduces less distortion to the original signal and sounds better. This decorrelation method is well suited for real-time applications for spatializtion, auralization, and upmixing.

In the current approach, the pole locations for the allpass filters are chosen in a naïve way and nothing prevents multiple decorrelation filters from having poles placed in the same locations. In the future, we would like to design the filters in a way that we can achieve the same or greater amounts of decorrelation with a shorter number of biquad sections by reducing the amount of competition between independent decorrelation kernels. Furthermore, the high transient content of sound files like the castanet recording expose the necessity of transient detection in decorrelation algorithms.

6. REFERENCES

[1] Garry Kendall, "The effects of multi-channel signal decorrelation in audio reproduction," in *Proceedings of the International Computer Music Conference*, 1994.

[2] Garry Kendall, "The decorrelation of audio signals and its impact of spatial imagery," *Computer Music Journal*, vol. 19, no. 4, pp. 71–87, Winter 1995.

[3] Maurice Boueri and Chris Kyirakakis, "Audio signal decorrelation based on a critical band approach," in *Proceedings of the 117 Audio Engineering Society Convention*, 2004.

[4] Ross Penniman, "A general-purpose decorrelation algorithm with transient fidelity," in *Proceedings of the 137th Audio Engineering Society Convention*, 2014.

[5] Guillaume Potard and Ian Burnett, "Decorrelation techniques for the rendering of apparent sound source widths in 3d audio displays," in *Proceedings of the 7th International Conference on Digital Audio Effects*, 2004.

[6] Andres Cabrera and Garry Kendall, "Multichannel control of spatial extent through sinusoidal partial modulation (spm)," in *Proceedings of the Sound and Music Computing Conference*, 2012.

[7] Andre Cabrera, *Control of Source Width in Multichannel Reproduction Through Sinusoidal Modeling*, Ph.D. thesis, Queen's University Belfast, 2012.
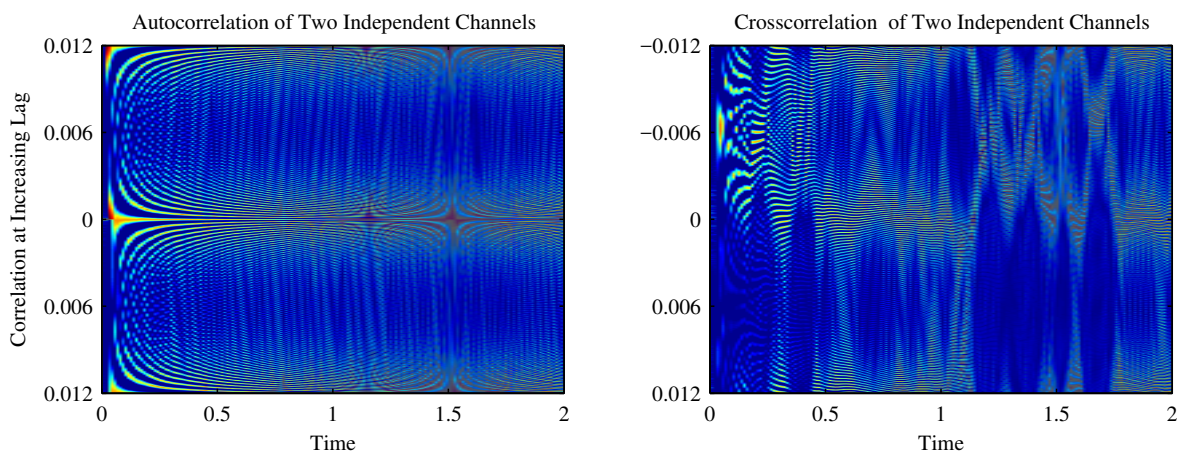
(a) *1024 allpass biquad sections*
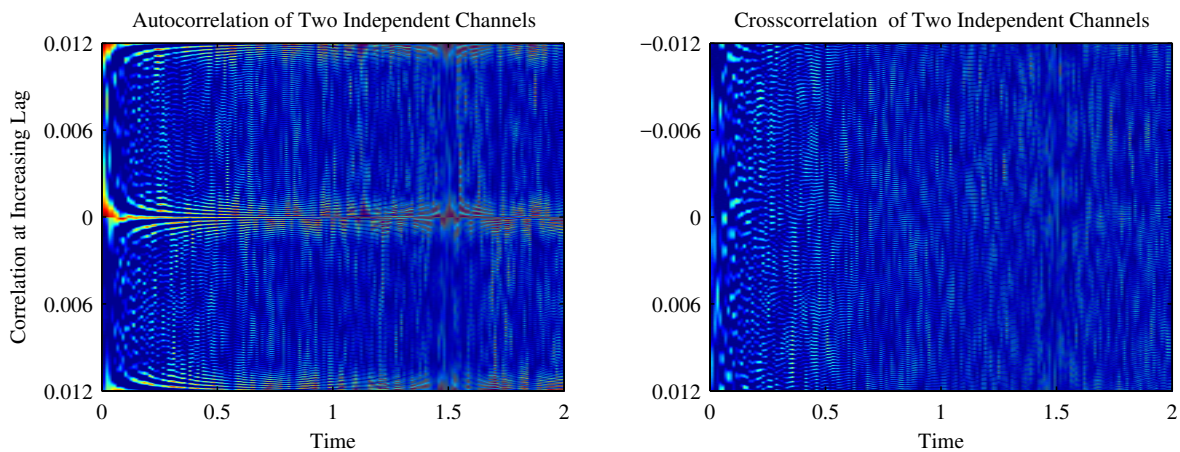
(b) *1024 sample length noise sequence*
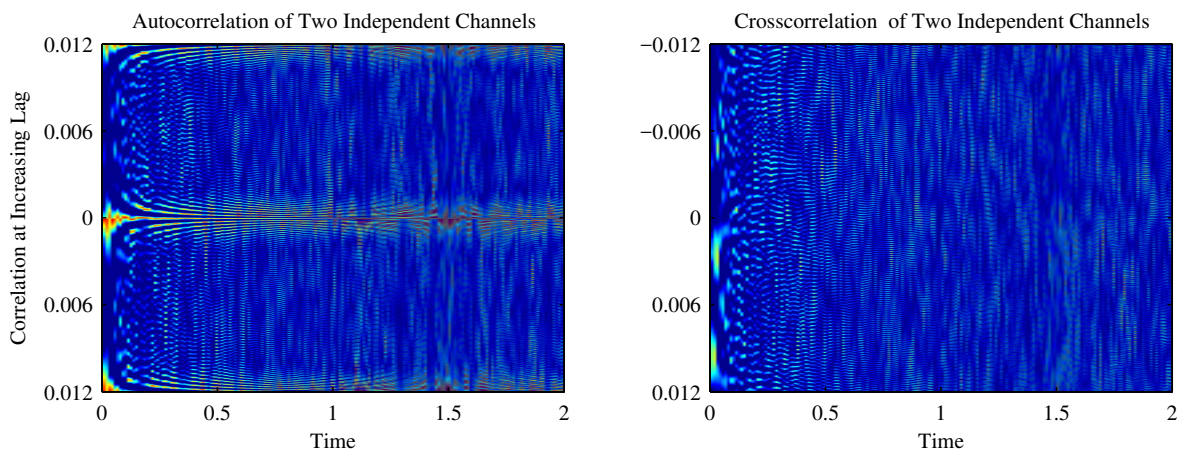
(c) *1024 sample length Kendall sequence*

Figure 6: *Auto-correlelograms (left) and cross-correlelograms (right) for* 20 Hz–20 kHz *linear sinusoidal chirp signal processed with allpass filter cascades (6a), noise sequences (6b), and Kendall filters (6c).*

[8] Mark Gardner, "Image fusion, broadening, and displacement in sound location," *Journal of the Acoustical Society of America*, vol. 46, no. 339, pp. 339–349, January 1969.

[9] Manfred Schroeder, "Natural sounding artificial reverberation," *Journal of the Audio Engineering Society*, vol. 10, no. 3, pp. 219–223, 1962.

[10] Frank Baumgarte and Christof Faller, "Binaural cue coding part i: Psychoacoustic fundamentals and design principles," *IEEE Transactions on Speech and Audio Processing*, vol. 11, no. 6, pp. 509–519, November 2003.

[11] Frank Baumgarte and Christof Faller, "Binaural cue coding part ii: Psychoacoustic fundamentals and design principles," *IEEE Transactions on Speech and Audio Processing*, vol. 11, no. 6, pp. 520–531, November 2003.

[12] Jonas Engdegård, Heiko Purnhagen, Jonas Rödén, and Lars Liljeryd, "Synthetic ambience in parametric stereo coding," in *Proceedings of the 116th Audio Engineering Society Convention*, May 2004.

[13] Jürgen Herre, Kristofer Kjörling, Jeroen Breebaart, Christof Faller, Sascha Disch, Heiko Purnhagen, Jeroen Koppens, Johannes Hilpert, Jonas Rödén, Werner Oomen, Karsten Linzmeier, and Kok Seng Chong, "Mpeg surround: The iso/mpeg standard for efficient and compatible multichannel audio coding," *Journal of the Audio Engineering Society*, vol. 11, no. 56, pp. 932–955, November 2008.

[14] Jean-Marc Valin, "Channel decorrelation for stereo acoustic echo cancellation in high-quality audio communication," in *Proceedings of Workshop on the Internet, Telecommunications and Signal Processing*, 2006.

[15] Franz Zotter, Matthias Frank, Georgios Marentakis, and Alois Sontacchi, "Phantom source widening with deterministic frequency dependent time delays," in *Proceedings of the 14th International Conference on Digital Audio Effects*, 2011.

[16] Sebastian Kraft and Udo Zölzer, "Stereo signal separation and upmixing by mid-side decomposition in the frequency-domain," in *Proceedings of the 18th International Conference on Digital Audio Effects*, 2015.

[17] Achim Kuntz, Sascha Disch, Tom Bäckström, and Julien Robilliard, "The transient steering decorrelator tool in the upcoming mpeg unified speech and audio coding standard," in *Proceedings of the 131st Audio Engineering Society Convention*, Oct 2011.

[18] Mathieu Lagrange, Luis Gustavo Martins, and George Tzanetakis, "Semi-automatic mono to stereo up-mixing using sound source formation," in *Proceedings of the 122nd Audio Engineering Society Convention*, May 2007.

[19] Robert Orban, "A rational technique for synthesizing pseudo-stereo from monophonic sources," *Journal of the Audio Engineering Society*, vol. 18, no. 2, pp. 157–164, 1970.

[20] Robert Orban, "Further thoughts on "a rational technique for synthesizing pseudo-stereo from monophonic sources"," *Journal of the Audio Engineering Society*, vol. 18, no. 4, pp. 443, 444, 1970.

[21] Jens Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization.*, The MIT Press, Cambridge MA, revised edition edition, 1996.
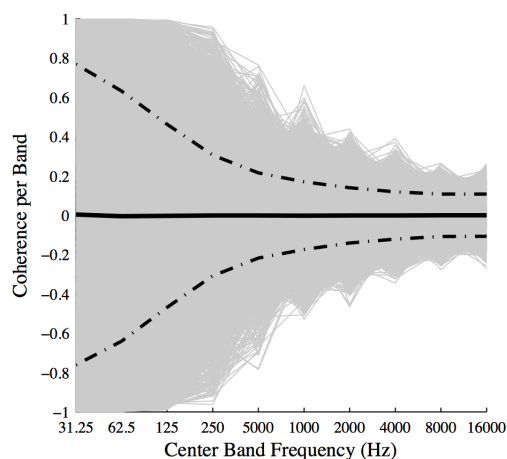
[22] "Ebu tech 3253 companion cd sound quality assessment material recordings for subjective tests (sqam)," https://tech.ebu.ch/publications/sqamcd, October 2008.
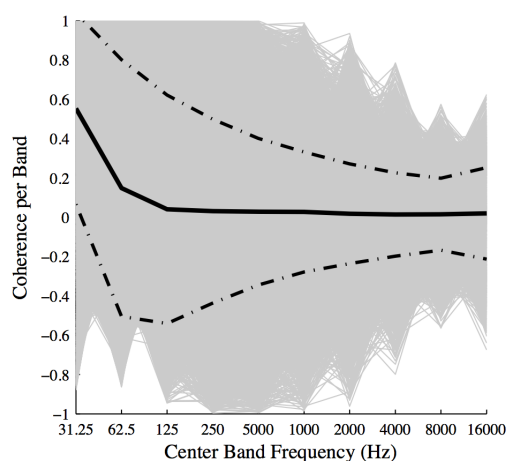
(a) *1024 allpass biquad sections*
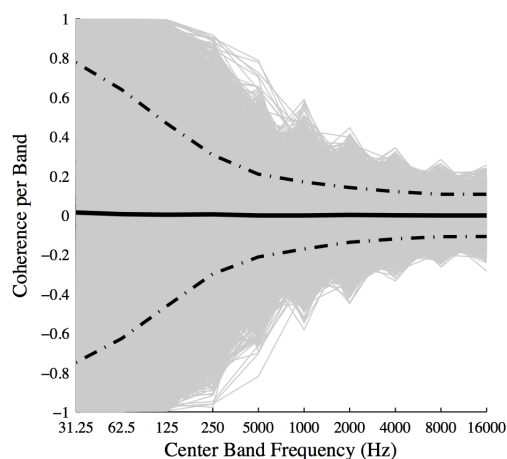
(b) *1024 sample length noise sequence*

(c) *1024 sample length Kendall sequence*

Figure 7: *Coherence in octave bands for allpass cascade (7a), noise sequence (7b), and Kendall filter (7c). The dark lines show the means, dotted lines the standard deviation, and gray lines individual one-vs-one coherences (1000 pairs).*