MULTICHANNEL CONTROL OF SPATIAL EXTENT THROUGH SINUSOIDAL PARTIAL MODULATION (SPM)†

Andrés Cabrera
Media Arts and Technology
University of California
Santa Barbara, USA
andres@mat.ucsb.edu

Gary Kendall
Artillerigatan 40
Stockholm, Sweden
garyskendall@me.com

ABSTRACT

This paper describes a new sound processing technique to control perceived spatial extent in multichannel reproduction through artificial decorrelation. The technique produces multiple decorrelated copies of a sound signal, which when played back over a multichannel system, produce a sound image that is spatially enlarged. Decorrelation is achieved through random modulation of the time-varying sinusoidal components of the original signal's spectrum extracted using a modified version of the Loris sinusoidal modeling technique. Sinusoidal partial modulation (SPM) can be applied in varying measure to both frequency and amplitude. The amount of decorrelation between channels can be controlled through adjusting the inter-channel coherency of the modulators, thus enabling control of spatial extent. The SPM algorithm has lent itself to the creation of an application simple enough for general users, which also provides complete control of all processing parameters when needed. SPM provides a new method for control of spatial extent in multichannel sound design and electroacoustic composition.

1. INTRODUCTION

Multichannel reproduction poses challenges to sound designers and electroacoustic composers that do not exist in traditional stereo. In particular, how does one control the listener's perception of spatial imagery across an expanded reproduction space, especially attributes like spatial extent. When listening to sounds in the real world, it is often easy to judge the size and extent of a sonic event. For example, the auditory image of a truck is larger, not only louder, than a cell phone, both of which appear smaller than the sound of the city in the background. If these three sounds were to be recorded and played back over a single loudspeaker, they would no longer be differentiated by the size of their auditory images. Most importantly, the background sound of the city would no longer be surrounding the listener.

The most straightforward idea for controlling spatial extent is to spread a signal across multiple loudspeakers, but this fails almost completely due to the influence of the precedence effect [2]. Controlling the relative distribution of amplitude across loudspeakers, as provided variously by VBAP [3], DBAP**,** and changing the order of Ambisonics [4], does nothing to address the influence of precedence, which varies with the source material and the relative size of the reproduction setting [2]. What can have an effect is the interaction of the loudspeaker signals with the acoustics of the room, but changes in spatial extent are rather like side effects. Wavefield Synthesis [5] can reconstruct complete acoustic soundfields, but provides no methodology for the control of perceived spatial extent.

The work presented here aims to provide a practical tool for controlling spatial extent in multichannel settings, from 5.1 and octophonic systems to three-dimensional loudspeaker arrays. Audio source material is manipulated to produce multiple decorrelated copies of a sound signal for distribution over a multichannel system. Decorrelation is achieved through random modulation of the time-varying sinusoidal components of the original signal's spectrum, extracted employing sinusoidal modeling. Additionally, by employing parameters outside their normative range, this technique can also be used for unusual creative sound processing.

2. BACKGROUND

2.1 Auditory Spatial Impression

Auditory Spatial Impression (ASI) is the characteristic of human auditory sensation associated with the acoustics of sources in a physical space. It attempts to group together all the sensations related to the spatial qualities and characteristics of the perceived sound. It has been described as composed of three distinct components: "spaciousness", "size impression," and reverberation [6]. It is generally accepted that "spaciousness" itself consists of at least two separate and distinct components [7,8]:

1. Apparent Source Width (ASW) is defined as the "width of a sound image fused temporally and spatially with the direct sound image
2. Listener envelopment (LEV) is defined as "the degree of fullness of sound images around the listener".

† This paper is based on material presented in [1].

ASW includes the sensations of broadness, blurriness and ambiguity of localization, while LEV imparts the sensation of fullness and surrounding [9]. ASW has also been called by some authors *perceived spatial extent* [10] and *individual source width* [11] in the context of loudspeaker reproduction.

An alternative approach to describing spatial impression has been proposed by Griesinger, who from the perceptual perspective of a recordist partitioned spatial impression into three components Continuous Spatial Impression (CSI), Early spatial impression (ESI) and Background spatial impression (BSI) [12]. These concepts and terminology, however, are not as widely used and cited as the previous.

The concept of Apparent Source Width is well accepted within the acoustics community, and it has been shown that the phenomenon is related the perception of spatial extent that occurs in multi-channel reproduction of incoherent or decorrelated signals [13]. However, to distinguish it from the acoustic phenomenon, the term "spatial extent" will be employed here for the phenomenon experienced in loudspeaker reproduction.

2.2 Incoherent Signals in Reproduction

The similarity of signals played back over multiple loudspeakers is instrumental in the perceptual fusion of these signals into a single auditory image in the phenomenon known as the *precedence effect*. However, if the signals are different, they will be perceived as separate sources originating from separate spatial locations. A particular and important case occurs when signals contain the same spectral components and energy distribution, but differ in their on-going phase relationships. Their time domain representations can be so different that there is no coherent temporal relationship between them. The simplest example is two independent noise signals, which will have the same spectrum but a wholly different and unrelated time-domain waveforms. When incoherent and spectrally identical signals are played back over two loudspeakers, the spatial image can vary from two identical sounding sources in two locations to one image with a broad spatial extent. There are three parameters that have been shown to influence the perception of spatial extent of the sound. They are:

1. The location of loudspeakers with respect to the listener and each other.
2. The amount of decorrelation between the signals in the speakers and its corresponding effect on the decorrelation between the signals at each ear.
3. The level difference between the loudspeakers.

As shown by Damaske [14], the broad spatial extent produced by incoherent noise is very clear when the loudspeakers are separated by a narrow angle. When the angle becomes wider, the image tends to dissociate and will be split between both loudspeakers, with less sound material perceived in between. For example, according to Damaske an angle of 90° between two frontal loudspeakers can result in dissociated and independent images. Damaske also investigated the effect of varying the amount of decorrelation in quadraphonic reproduction, and found that when the degree of incoherence for a band of noise increased, so did the perception of spatial extent. Wagener showed that the degree of envelopment could also be controlled with relative signal levels using delayed incoherent reflections. As the level of the incoherent reflections increased, so did the perceived envelopment [14].

2.3 Fluctuations of ITD and ILD

It has also been shown that modulation of interaural time difference (ITD) and interaural level difference (ILD) can affect the perception of spatial extent in a similar way. Aschoff showed that fast moving sources can produce wide images when the speed of movement is too fast for the auditory system to track [14]. Griesinger found that fluctuations of ITD and ILD with frequencies lower than 3Hz have the effect of making the perceived sound move [15]. However frequencies greater than this will result either in a wider image, or the perception of a narrow image in the presence of a surrounding ambiance. Mason et al has showed that the magnitude of fluctuations in the ITD is related to the perception of ASW [16].

Mason et al. later showed that the relation between decorrelation and ASW is mediated by frequency, as some frequency areas, like the mid range, require more decorrelation to be perceived as wide as lower frequencies with less decorrelation [17].

3. SINUSOIDAL PARTIAL MODULATION (SPM)

3.1 Rationale

There is a long history of audio techniques that aim to enhance perceived spatial extent. The vast majority of these techniques involve manipulations in the time domain, though many produce spectral artifacts such as coloration and phasiness. Of these, particularly noteworthy is that of using random-phase all-pass filters first proposed by Kendall [10], because the technique is able to produce controlled levels of decorrelation among multiple audio channels. These decorrelation filters were shown to affect the precedence effect as well as perceived source width. Potard and Burnett [18] enhanced Kendall's all-pass filtering technique by decomposing the sound into three sub-bands (Low 0-1 kHz, Mid 1-4 kHz and High 4-20 kHz) that enabled different amounts of decorrelation to be applied. Both techniques suffer artifacts due to their static filtering, because the localization cues for any particular band of frequencies is static.

Discussed here is an innovative technique especially appropriate to multi-channel reproduction called Sinusoidal Partial Modeling (SPM). This technique applies controllable amounts of dynamic modulation to the partials of a source signal. Through the technique of sinusoidal modeling a set of time-varying sinusoidal partials together with residual energy information can be extracted from a source signal. Decorrelation can then be introduced at the resynthesis stage through modulation of the frequency and amplitude of the partials, which are resynthesized using oscillator banks, one for each output channel. Thereby any number of decorrelated copies of

a monophonic source signal can be created. In multi-channel reproduction, they produce a wide image. This approach to decorrelation can offer the following advantages over previous techniques:

1. Any number of output channels can be produced.
2. Since there is no time-domain filtering involved and modulation can be kept below perceptual thresholds while still achieving decorrelation, only spatial characteristics of the source, i.e. source width, should be affected.
3. Because the decorrelation is dynamic, the typical artifacts of static decorrelation like phasiness or static location cues will not be present. The product of this technique can resemble natural spatial widening occurring due to reverberation because of this dynamic nature.
4. SPM can provide control over the source width for multichannel playback, as the decorrelation can be carefully tailored to different circumstances, by affecting parts of the spectrum in different ways or controlling the relation between decorrelation and speaker location
5. The algorithm can also be used as a creative sound design tool to modify the sound drastically through extreme modulation (beyond the point where it is clearly audible as pitch deviation) that will have spatial effect as it is different for every channel.

There are also some disadvantages to this approach. Sinusoidal modeling is applicable to processing a wide variety of sources although it tends to be less successful at capturing full mixes, which might include frequent complex transients and quickly varying noise and sinusoidal components. It will generally perform better with individual tracks or submixes which can be mixed together. However, in the present implementation, effort was made to minimize artifacts through custom improvements, so that full mixes are rendered more successfully given appropriate settings.

3.2 The Algorithm

Sinusoidal Modeling is well suited to allow independent modulation of each of the components with the knowledge that the perceived source identity will be preserved if the modulation stays within perceptually detectable thresholds. Additionally, the amount of decorrelation among different regions of the spectrum within a single channel or between channels can be precisely controlled through controlling the similarity between the modulating signals for the bands.

The algorithm presented here attempts to avoid the obvious drawbacks of sinusoidal modeling by adopting and enhancing the Loris model [19]. The Loris technique for analysis/resynthesis was chosen because it uses the time-frequency reassignment method, which produces greater precision in time and frequency for a particular window size than other frequency estimation techniques. This means the analysis can use smaller windows with better frequency resolution than regular FFT, while also reducing time and transient smearing through time reassignment. This allows for a very high precision of sinusoidal tracking while giving adequate transient representation. Additionally, Loris provides a method for representing the residual/stochastic energy of a signal in the form of energy "band-width," which is assigned to partials then recreated using "band-width enhanced" oscillators. Consequently, the inter-channel decorrelation level of the stochastic energy can be controlled as precisely as the deterministic part. These two characteristics make Loris a good starting point for the system discussed here, as it is best able to represent most types of practical signals.

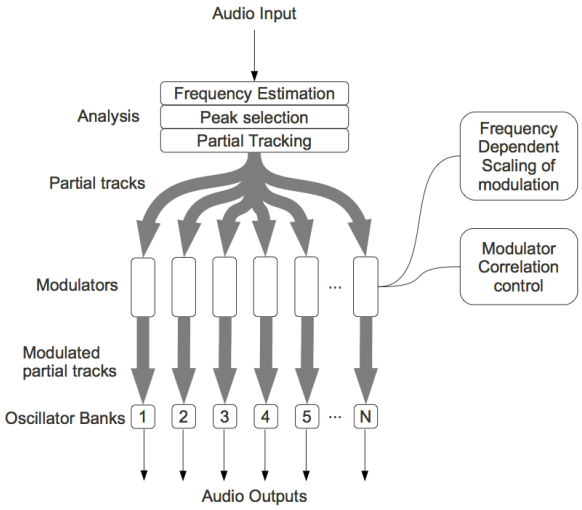

Figure 1. Overview of algorithm for Sinusoidal Partial Modulation.

3.3 Resynthesis and Decorrelation

A set of oscillator banks, one for each output channel, resynthesizes the input signal based on its sinusoidal analysis. As specified in Loris, the basis for resynthesis is the bandwidth-enhanced oscillator, which has frequency, amplitude and bandwidth as parameters for each breakpoint in the partial tracks. It is at this stage that the frequency and amplitude are modulated to produce inter-channel decorrelation. The modulator curves for partial track modulation need to have the following characteristics:

1. They must be random but low-passed to sub-audio frequencies to avoid audible sidebands.
2. The values must be clamped within +/- *max* to limit the maximum deviation from the original frequency and to have no DC offset.
3. They should be economical in terms of CPU usage, as a great number of them need to be calculated.
4. The signal must not make big sudden jumps that could easily stand out.

The random modulators could be constructed by low-pass filtering white noise, which would generate band-limited signals, but having that many filters running continuously would have a huge impact on CPU load. Additionally, although the range can be easily limited, there would be no way of preventing large jumps in the signal other than reducing the range. Because of this, an alternative simple method for generating low-passed modulator signals was developed. The signal is constructed by performing quadratic interpolation between random val-

ue breakpoints that are produced at regular intervals. How the random breakpoint generator works is shown in Figure 2. The maximum frequency of the modulator signal can be controlled and will be half the rate at which new breakpoints are produced. Each new random breakpoint is limited to a range around the previous one, in a sort of random walk algorithm. This guarantees that the jumps are never too large. Additionally to make sure the random values stay within the upper and lower limits, it is necessary to clamp the edges of the random range, forcing the breakpoint values to tend toward the center when they are close to the edges. This can be seen in Figure 2 at point 5, where the top of the range from point 4 has been reduced.

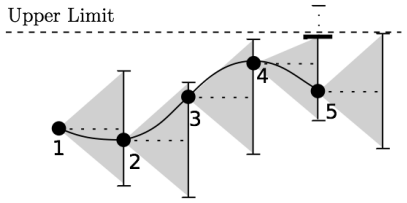

Figure 2. Random value generation constraints applied to one modulator signal.

The starting "phase" of the random modulator update function are randomized, so that each modulator will produce points at a different moments in time. This is necessary to avoid having high likelihood of the modulators peaking and having minimums at the same point in time, or other parallel motion that could be perceived.

An example of an actual modulator signal is shown in Figure 3, where random points are generated every 512 points, with a maximum jump of 1. Due to the interpolation, the modulator curves can occasionally and briefly cross the limits. This is not a problem as it would only mean a minor temporary increase in frequency or amplitude deviation.

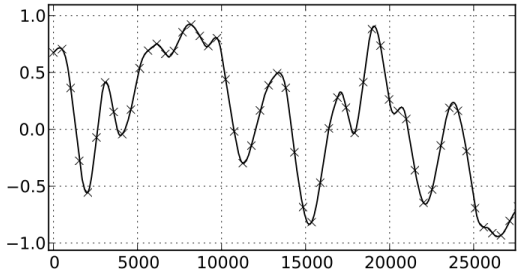

Figure 3. Resulting random modulation curve.

The coherency of the modulators and therefore the coherency between the output signals can be controlled by having a modulator bank which can be mixed together in any desired way to produce the final modulators. In this way coherency between output signals can be carefully controlled. Although not a widely studied subject, it seems likely that random modulation occurring within a critical band could cancel out or be diminished within the ear. To ensure this does not happen, one independent modulator is used per critical band (24 in total) so that sinusoidal partials falling within the same channel and same critical band are modulated by highly correlated signals.

3.4 SPM Implementation

The SPM algorithm was implemented in C++ to be able to reuse as much code from Loris as possible, though many parts of Loris still needed to be rewritten or modified. The resulting program is called Sprokit (SPatial PROcessing KIT) is designed for multiple platforms, including Linux, OS X and Windows. Any audio processing like Sprokit that involves FFT-based processing will introduce latency as a window of samples must be accumulated before any process can take place. Additionally, this process must accumulate the output of at least two windows in order to do the partial tracking as trajectories to the peaks of the second window start already during the resynthesis time of the first window.

Because the calculations for the modulators require more CPU time than current systems can provide, the program currently runs offline, that is, it must load a file and write the output files to disk, rather than streaming to an audio card. However, it is internally designed to eventually meet real-time requirements. It implements streaming analysis and resynthesis, which has the benefit for offline processing of enabling the processing to stop at any moment, while still producing a valid output audio file.

The interface has a main window showing the most important parameters that affect the spatial properties of the sound like relative level of the output signals and amount of decorrelation. The rest of the parameters are in a separate "properties" dialog window which is available if the user requires more advanced control. Figure 4 shows the final graphic user interface running on Linux. A set of sliders to allow per-channel adjustment of level (trim) and wet/dry mix. This was deemed useful as level and mix enable an engineer or composer to adjust the spatial image, a clearly desirable and practical feature. This particular implementation is limited to eight output channels, which was considered sufficient for most practical uses, although the algorithm itself has no such enforced limit.

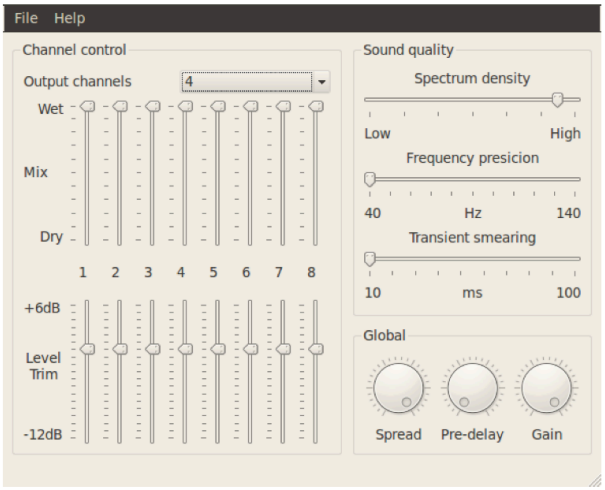

Figure 4. Main graphical user interface for Sprokit.

The algorithm parameters not presented in the main window of the application are accessible in the properties dialog window from the application's menus. This dialog window, has an initial page containing the parameters for analysis and resynthesis, and two additional pages with the specific parameters for frequency and amplitude modulation.

A simple preset mechanism was developed to be able to quickly switch between different configurations for consistent testing. Separate presets for the analysis and the resynthesis parameters are implemented to allow holding one constant while experimenting with the results of the other.

3.5 Objective Evaluation of SPM

A comparative objective study was performed to determine the effects of the SPM technique on interaural cross correlation (IACC) and other computational measures predicting spatial extent. The tests were conducted to verify that the algorithm produced interaural effects in a similar way to other techniques known to produce an enlarged source width, such as incoherent noise or sound processed through all-pass random-phase filters. To perform the tests, a set of typical audio signals were prepared in addition to coherent and incoherent noise. These signals were passed through the Kendall decorrelation algorithm [10] and the SPM multichannel decorrelation. Then, all the processed and original signals were rendered binaurally using convolution of HRIRs positioning them according to five selected loudspeaker configurations and simulating the effect of the cross-talk that would occur if the signals were played back over loudspeakers in the specified locations. These binaural signals were evaluated using three different metrics (Interaural cross-correlation coefficient, Interaural cross-correlation fluctuation function and Mason's Perceptually Motivated Measurement of Spatial Sound Attributes (PMMP)) to verify that the algorithm affects predictors of source width. The result confirmed that SPM affected predictors of spatial extent as expected and in ways comparable to Kendall decorrelation.

3.6 Strategies for Multichannel Sound Design

A wide image width is obtained when decorrelated signals are routed to each separate loudspeaker. The most continuous impression of width is likely to be achieved when loudspeakers are not too far apart; otherwise, there might be an empty space perceived between the speakers. Using a higher density of loudspeakers tends to produce a better impression of width and of being surrounded by the sound.

For a 5.1 setup, when the source is monophonic, the original source can be placed on the center speaker, with decorrelated copies on the other four. This creates a frontal bias, because transients will be stronger and sharper in the original signal, and the decorrelated copies will spread the sound around the listener. If a center loudspeaker is not available, the original sound can be mixed into the front channels. Using pre-delay in this case can help the center channel stand out, or blend better if desired.

A useful strategy when processing a stereo sound source for an octophonic layout is to place the original stereo signal in the front pair of speakers without transformation (dry only), and then on the other three speakers on the left, decorrelated copies of the left channel, and similarly for the right. This has the effect of preserving the original transients of the signal, with strong bias towards front localization, while still producing an effective spread of the sound in all directions. The lateral (left-right) separation of the original stereo source is also accurately preserved.

Since the processing works best for individual sources, it can be used as part of the compositional process to treat individual elements independently. The decorrelated copies of elements can then be positioned and used as desired, allowing for different width parameters, spread and spatial locations and for each source. The processed and unprocessed signals can be distributed across space by treating them as independent objects using techniques like VBAP or ambisonics. This would allow moving sources as well as potentially interesting spatial effects like adjusting source width dynamically or merging and separating sources.

4. CONCLUSIONS

This paper has presented a novel method for controlling apparent spatial extent in multichannel reproduction. The Sinusoidal Partial Modulation (SPM) method produces and controls the multichannel decorrelation of audio signals, through bringing together in an unusual way two usually separate areas of audio signal processing: sinusoidal modeling and decorrelation. The SPM algorithm is suitable for processing as wide a variety of audio material as sinusoidal modeling is. For the most satisfying results, the algorithm should be applied to individual tracks or submixes.

Most existing decorrelation techniques are based either on static phase shifts (generally using time domain filtering) or on applying modulation across the time domain signal as a whole. The SPM decorrelation technique is in a way a mixture between the two since the amount of phase shift varies through random modulation and is different for different areas of the spectrum. Thus, the decorrelation produced by this technique is both frequency dependent and dynamic. The SPM technique also allows very fine control over the decorrelation in relation to frequency. This opens new possibilities for the exploration of the frequency dependence of decorrelation. Then too, dynamic decorrelation is less prone to the kind of artifacts typical of other techniques. For example, the timbre of the output appears less colored.

Finally, the algorithm lent itself to the creation of an application simple enough for general users, but providing complete control of all processing parameters when needed.

5. REFERENCES

[1] A. Cabrera, *Control of Source Width in Multichannel Reproduction Through Sinusoidal Modeling*, Ph.D. dissertation, Queen's University Belfast, 2012.

[2] G. Kendall and A. Cabrera, "Why Things Don't Work: What You Need To Know About Spatial Audio," *Proceedings of the International Computer Music Conference,* Huddersfield, England, 2011.

[3] V. Pulkki, "Virtual sound source positioning using vector base amplitude panning," *J. Audio Eng. Soc.,* vol. 45, no. 6, pp. 456–466, 1997.

[4] D. G. Malham and A. Myatt, "3-d sound spatializationambisonic techniques," *Computer Music Journal,* vol. 19, no. 4, pp. 58–70, 1995.

[5] M. M. Boone and E. N. Verheijen, "Multichannel sound reproduction based on wavefield synthesis," in *Audio Engineering Society Convention 95*, 1993.

[6] T. Okano, L. Beranek, and T. Hidaka, "Relations among interaural cross-correlation coefficient (IACCE), lateral fraction (LFE), and apparent source width (ASW) in concert halls," *J. Acoust. Soc. Am.,* vol. 104, no. 1, pp. 255–265, July 1998.

[7] M. Morimoto and Z. Maekawa, "Effects of low frequency components on auditory spaciousness," *Acustica*, vol. 66, pp. 190–196, 1988.

[8] J. S. Bradley and G. A. Soulodre, "Objective measures of listener envelopment," J. *Acous. Soc. Am.*, vol. 98, no. 5, pp. 2590–2597, November 1998.

[9] A. M. Sarroff and J. P. Bello, "Toward a computational model of perceived spaciousness in recorded music," *J. Audio Eng. Soc.,* vol. 59, no. 7/8, pp. 498–513, 2011.

[10] G. Kendall, "The decorrelation of audio signals an its impact on spatial imagery," *Computer Music Journal*, vol. 19:4, pp. 71–87, 1995.

[11] F. Rumsey, "Spatial quality evaluation for reproduced sound: Terminology, meaning, and a scene-based paradigm," *J. Audio Eng. Soc*., vol. 50, no. 9, pp. 651– 666, September 2002.

[12] D. Griesinger, "The psychoacoustics of apparent source width, spaciousness and envelopment in performance spaces," *Acta Acustica*, vol. 83, no. 4, pp. 721–, July/August 1997.

[13] R. Mason and F. Rumsey, "A comparison of objective measurements for predicting selected subjective spatial attributes," in *112th AES Convention*, May 1013 2002.

[14] J. Blauert, *Spatial hearing: the psychophysics of human sound localization*, revised ed. MIT press, 1997.

[15] D. Griesinger, "The psychoacoustics of Apparent Source Width, spaciousness & envelopment in performance spaces," Lexicon, Tech. Rep., 1997.

[16] R. Mason, F. Rumsey, and B. de Bruyn, "An investigation of interaural time difference fluctuations, Part 1: the subjective spatial effect of fluctuations delivered over headphones," in *110th AES Convention*, 2001.

[17] R. Mason, T. Brookes, and F. Rumsey, "Frequency dependency of the relationship between perceived auditory source width and the interaural cross-correlation coefficient for time-invariant stimuli," *J. Acous. Soc. Am.,* vol. 117, no. 3, pp. 1337–1350, 2005.

[18] G. Potard and I. Burnett, "Control and measurement of apparent sound source width and its applications to sonification and virtual auditory displays," in *Proceedings of ICAD 04-Tenth Meeting of the International Conference on Auditory Display*, 2004.

[19] K. Fitz and L. Haken, "On the use of time-frequency reassignment in additive sound modeling," *J. Audio Eng. Soc*, vol. 50, no. 11, pp. 879–893, 2002.