Signal-Aware Spatial Positioning

Christopher Latina
Georgia Tech Center for Music Technology
840 McMillan St. Atlanta GA 30332
Chris.latina@gatech.edu

Alexander Lerch
Georgia Tech Center for Music Technology
840 McMillan St. Atlanta GA 30332
alexander.lerch@gatech.edu

ABSTRACT

The currently available tools for the spatial composition of music are limited in multiple ways. Simple panning tools or mixer-based manipulation techniques are often tedious to use and do not scale to different loudspeaker configurations. This paper presents a signal-aware spatial positioning technique that extends the functionality of these traditional approaches. Using spectral and rhythmic features from the input, the movement and positioning of sound in space can be automated and is both signal-adaptive and user-interactive. The employment of ambisonics allows for easy scalability. A software prototype illustrates an interface for composition, performance, or installation as a basis for further exploration within the domain.

Author Keywords

NIME, Spatial Composition, Signal-aware, Ambisonics, ATK, SuperCollider

ACM Classification

H.5.5 [Information Interfaces and Presentation] Sound and Music Computing
H.5.2 [Information Interfaces and Presentation] User Interfaces
H.5.1 [Information Interfaces and Presentation] Multimedia Information Systems

1. INTRODUCTION

Multichannel spatialization can create complex movement of sound and add an additional compositional dimension. John Chowning's 1972 piece titled *Turenas* explored spatialization within a quadraphonic setup [1]. Chowning discusses using angular cues rather than channel mixing to move sound [2]. V. Pulkki's Vector base amplitude panning extends Chowning's panning control function approach for pair-wise and triplet-wise amplitude panning [3]. These ideas inspired the following objectives which led research methodology and software programming decisions. The focus of this work was on creating a deliverable software prototype for spatial composition.

- Design a multichannel and real-time panning system that uses modulation signals or waveforms to move sound in space independent of the loudspeaker configuration.
- Explore meaningful audio feature-based mappings to modulate spatial position of signal.
- Design a method for the user to control the effect of the audio features on spatial positioning and higher-level automation parameters.
- Design and implement an interface for performance, composition, and installation (automation).

2. RELATED WORK

2.1 Immersive Sound

While surround sound has become ubiquitous within the fields of music technology, cinema, installation art, and immersive theatre, many of the widely available tools for creating multichannel compositions are channel-specific or constructed for a standardized system or a single performance setting. A number of different approaches for multichannel sound reproduction exist within a variety of commercial industries. While 5.1 surround sound systems have become popular even in home theater systems, more specialized systems provide new immersive experiences.

Although ambisonics has existed since the 1970s under research for the British NRDC, only recently has a resurgence of interest begun. BBC's R&D team has published a series of white papers covering the subjects of ambisonics in the context of broadcast in 2012 [4] and elevated spatial source localization with higher order ambisonics in 2013 [5].

Dolby has also recently incorporated ambisonics into their research and product development. Dolby released a new product called *Atmos*, which immerses the audience in detailed moving sound. In 2012, Dolby acquired Barcelona-based company *imm sound*, which specialized in higher-order ambisonics.

Wave field synthesis uses arrays of speakers to generate a holophonic audial image known as a wave front [6]. This system is unique in that it approximates virtual sound sources that exist in front of or behind the plane of speakers [7]. An example of this system is the Game of Life Foundation, which hosts a number of composers to write for their 192-speaker system [8].

4DSOUND is a new immersive environment comprised of 16 columns with 3 sets of speakers in each. Underneath the floor, 9 sub speakers project upwards to the audience standing above [9]. Collaborating with Ableton, their software allows the performer or composer to draw the movement using a Max for Live and an iPad application.

Lastly, IRCAM's OMPrisma for the OpenMusic environment enables high level compositional control over spatial positioning of sound in conjunction with synthesis [10]. One example in Shumacher's and Bresson's paper describes using audio files as control signals to vary azimuth from center position determined by spectral features and onset detection. This approach however appears to be strictly for composition and not real-time performance.

2.2 Signal-Aware Control

The use of audio-features to control sound localization offers a fundamentally new approach to spatial composition. With further development and standardization of ambisonic-like approaches to spatial audio in the commercial field, the demand for new compositional interfaces may rise. While this approach is novel, it offers power automation that uses direct audio input, providing modulation for spatial transformations nescient of synthesis techniques or score. Signal-aware control allows for real-time and interactive use in conjunction with automation. The result generates semantically viable mappings with dynamic range of control.

3. SOFTWARE

3.1 Design

SuperCollider supports a number of existing unit generators for simple stereo panning and extensions for vector base amplitude panning (VBAP)[1] and the Ambisonics Toolkit (ATK). The Ambisonics Toolkit, developed by Joseph Anderson, provides an extensive number of first order transformations [11]. These transformations are based on virtual microphone polarity patterns.

ATK's transformations also mimic the localization cues described in Chowning's paper, factoring in inter-aural time delay differences in sound pressure level to create a realistic audio image. The toolkit allows for spatial-centric composition to scale to a number of speaker configurations independent of the number of speakers or output channels.

Furthermore, the Machine Listening library in SuperCollider provides a set of tools to analyze input audio buffer streams. The output of the machine listening unit generators are audio features that can be used as control signals [12] such as modulation sources for (in this case) spatial simulation cues [2]. The subset of audio features selected for the current software implementation comprise of the spectral centroid, the spectral flatness, the spectral percentile, as well as detected onsets as modulation sources for controlling the rotational ambisonic first order transformation.

The application uses ATK's pantophonic (2D) regular polygon decoder and rotational transformation. Given a mono or stereo audio input and a number of channels, the virtual speaker encoder assumes an evenly distributed speaker configuration. A 3-channel setup assumes an equilateral triangle; a 4-channel setup assumes a square, quadraphonic setup, etc.

3.2 Interface Components

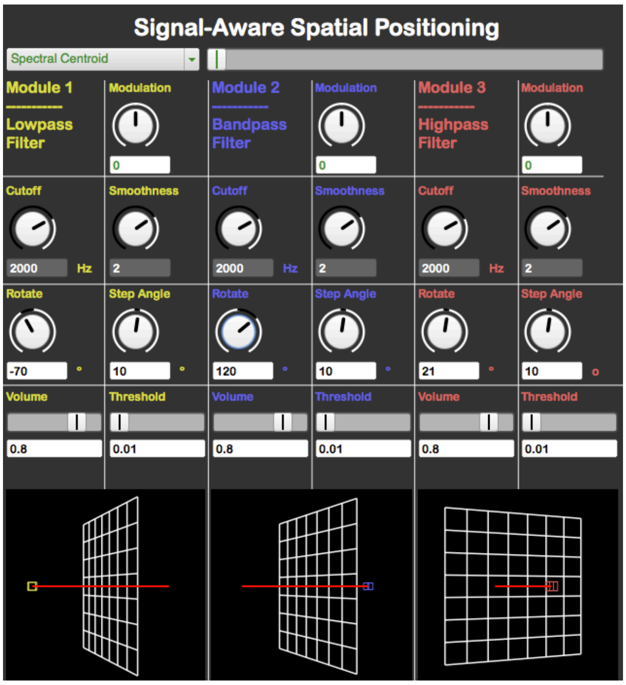

Figure 2. Interface for SuperCollider software

Figure 2 displays a screenshot of the interface with its controls. At the top, the user selects an audio feature from the dropdown; the options available for selection include *spectral centroid*, *spectral flatness*, and *spectral roll-off*.

The input signal is filtered into three discrete bands: a *lowpass filtered* band for the lower register, a *bandpass filtered* band for the middle register, and a *highpass filtered* band for the upper register. The application's graphical interface is split into these three main vertical modules which correspond to the three filter bands. The intention of filtering the input signal is to allow more complex rotations, such as positioning low register spectral content slowly in clockwise motion while high register spectral content moves quickly in a counter-clockwise manner. The *Cutoff* knob within each band controls the filter's cutoff frequency.

The *Modulation* knob shown in Figure 3 controls the modulation amount and direction of the spectral feature toward rotating the respective band's sound field. The knob is initially centered at 0 to provide directionality. A negative modulation index rotates the sound field counter clockwise. If the spectral centroid increases, the sound field rotates further in the direction specified.

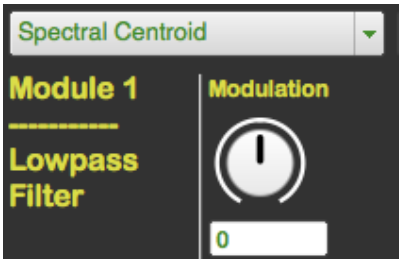

Figure 3. Feature Modulation

The *Step Angle* control rotates the sound field by the given amount upon detection of an onset within the filtered band. The range for this control knob is +/- 180 degrees. If *Step Angle* is set to +30, the filtered sound field will rotate clockwise by 30 degrees upon detection. Threshold controls the threshold at which an onset is detected.

Smoothness controls a filter that smooths the modulation signal routed from the output of the onset detection and the spectral feature output multiplied by the *Modulation* amount. A high level of smoothness creates fluid transformations while reducing smoothness allows for jumps and jitters.

Rotate is a manual control for rotating the band's sound field by +/- 360 degrees. *Volume* is a also a manual control. The 3D canvas displays a grid and an arrow that animates the computer rotational angle for each band.

3.3 Internal Program Structure

3.3.1 The SynthDef

The majority of the feature-mapping techniques are calculated within a synth definition in the SuperCollider script. Figure 4 demonstrates the program flow of audio analysis and spatial mapping. The spectral modulation mapping is performed on the unfiltered audio input signal. This can be a pre-recorded buffer or line in routed from an audio interface. The synthDef is instantiated for each band with a parameter to specify the filter type (lowpass, bandpass, and highpass). Onset detection and subsequent mappings are performed after filtering. An impulse control rate is sent at a rate of 30 Hz (or 30 frames per second) to update the 3D canvas visualization via Open Sound Control (OSC).

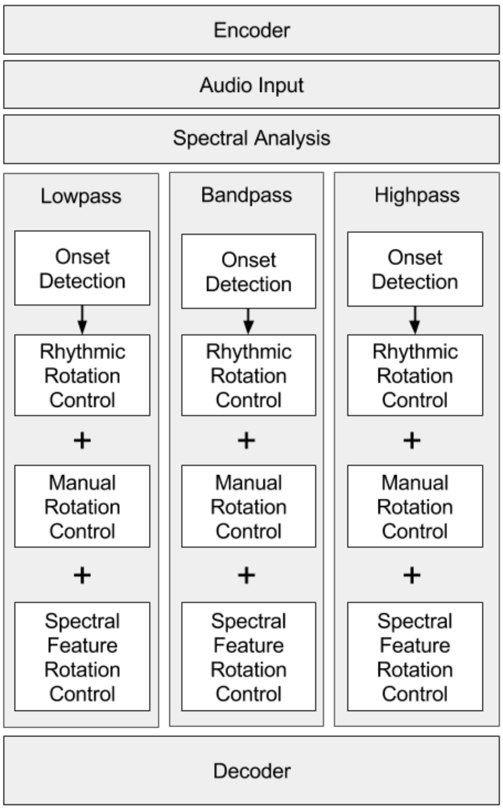

Figure 4. **Internal Program Flow**

4. CONCLUSION

We presented a prototype as an initial proof of concept for signal-aware spatial positioning of audio. This approach is cognizant of the audio input, which creates automated mappings that correspond to audio features without the need for knowledge of synthesis metadata or score. Therefore, spatial position is decoupled from audio synthesis but still reflects the timbral content of the audio signal. Further extension of the modulation techniques presented can provide a composer with complex routings for control and automation of form.

4.1 Additional Transformations

Implementation of additional ambisonics transforms, especially those in 3D space will further the complexity of the tool. Tilt and Tumble, for example, are the equivalent first order B-format rotational transforms around the X and Y-axes. [11]

4.2 Additional Signal Features

While onset and spectral features are the internally specified mappings to trigger rotation, additional control signals can be used. The Machine Listening library, for example, also contains a beat tracking function that deciphers tempo, pitch following, key tracking, and loudness [12]. Logical applications of these features would require further thought. While beat tracking may correspond with the Step Angle onset rotational approach, pitch following may logically correspond with a chromatic circle of pitches mapped to the virtual speaker configuration.

4.3 Multiple Input Signals and Modular Routing

Scaling the class of functionality for multiple inputs and modular routing would significantly improve the prospects for this tool. Additional instantiations could be used to route live or pre-recorded multi-track signals for spatialization. Rather than filtering a "mixed" signal, individual tracks for each instrument could be controlled. Furthermore, side-chaining one signal's features to modulate another signal's position could facilitate interaction. Abstracting the control interface and extending Open Sound Control would also significantly extend the tool's ability. Sending feature control signals as OSC messages to external devices/effects or visa versa responding to OSC messages from a DAW would allow motion sequence control.

4.4 Extension

Further use of MIR and feature motivated compositional and sound engineering tools have yet to be explored. Feature extraction from audio signals allow for meta-information to be mapped to generate additional compositional elements, such as spatial positioning in this case. The approach is powerful in that extraction allows for new mappings for composition audio presentation without existing knowledge of synthesis or score. This concept can be applied to the automation of controlling effect module parameters in the signal chain or generating parts for additional instruments, while still providing user-interactivity for live performance.

5. ACKNOWLEDGEMENT

The author wishes to express his appreciation to Chi-wei Wu for his insightful suggestions.

6. REFERENCES

[1] M. Meneghini. Stria, by John Chowning: Analysis of the Compositional Process. *Proceedings of the XIV Colloquium on Musical Informatics (XIV CIM 2003).* Firenze, Italy, 2003.

[2] J. Chowning. The Simulation of Moving Sound Sources, *J. Audio Eng. Soc.*, 19 (1). 2-6.

[3] V. Pulkki. Compensating Displacement of Amplitude-panned Virtual Sources. *Audio Engineering Society 22th Int. Conf. on Virtual, Synthetic and Entertainment Audio*, (2002), 186-195.

[4] C. Baume, A. Churnside, Upping the Auntie: A Broadcaster's Take on Ambisonics, *BBC R&D Publications: White Paper WHP 221*, 2012. http://www.bbc.co.uk/rd/publications/whitepaper22, accessed Jan 30, 2015

[5] P. Power, C. Dunn, W. Davies, and J. Hirst, Localisation of Elevated Sources in Higher-order Ambisonics, *BBC R&D Publications: White Paper WHP 261*, 2013. http://www.bbc.co.uk/rd/publications/whitepaper261, accessed Jan 30, 2015

[6] M. Baalman. Spatial Composition Techniques and Sound Spatialisation Technologies, *Organised Sound* 15(3), 209–218.

[7] L. Jung-Min, C. Jung-Woo, K. Yang-Hann. Wave Field Synthesis of a Virtual Source Located in Proximity to a Loudspeaker Array, *Journal Of The Acoustical Society Of America* 134, 2106-2117.

[8] The Game of Life. http://gameoflife.nl/en/, accessed December 4, 2014

[9] 4DSOUND. http://4dsound.net/overview/, accessed December 4, 2014

[10] M Schumacher and J. Bresson, Spatial Sound Synthesis in Computer-Aided Composition, *Organised Sound* 15(3), 271-289.

[11] J. Anderson. The Ambisonic Toolkit: tools for soundfield-kernel composition. *http://www.ambisonictoolkit.net/wiki/tiki-index.php?page=SC3+Help,* December, 2014.

[12] S. Wilson, N. Collins, and D. Cottle. *The SuperCollider Book*. Cambridge, Mass: MIT Press, 2011.

7. APENDIX

A repository of this code can be found at the following link: https://github.com/chrislatina/7100.

A video of the software can be found at the following link: http://youtu.be/Ljeivc7NF2I