SPECTER: COMBINING MUSIC INFORMATION RETRIEVAL WITH SOUND SPATIALIZATION

Bill Manaris  
Computer Science Dept.  
College of Charleston, USA  
manarisb@cofc.edu

Seth Stoudenmier  
Computer Science Dept.  
College of Charleston, USA  
stoudenmiersh@g.cofc.edu

ABSTRACT

Specter combines music information retrieval (MIR) with sound spatialization to provide a simple, yet versatile environment to experiment with sound spatialization for music composition and live performance. Through various interfaces and sensors, users may position sounds at arbitrary locations and trajectories in a three-dimensional plane. The system utilizes the JythonMusic environment for symbolic music processing, music information retrieval, and live audio manipulation. It also incorporates Iannix, a 3D graphical, open-source sequencer, for real-time generation, manipulation, and storing of sound trajectory scores. Finally, through Glaser, a sound manipulation instrument, Specter renders the various sounds in space. The system architecture supports different sound spatialization techniques including Ambisonics and Vector Based Amplitude Panning. Various interfaces are discussed, including a Kinect-based sensor system, a Leap-Motion-based hand-tracking interface, and a smartphone-based OSC controller. Finally, we present Migrant, a music composition, which utilizes and demonstrates Specter’s ability to combine MIR techniques with sound spatialization through inexpensive, minimal hardware.

1. INTRODUCTION

Speaker (spˈkər) n.
1. a visible incorporeal spirit, esp. one of a terrifying nature; ghost; phantom; apparition.
2. some object or source of terror or dread.
Also, esp. Brit., spectre.
[1595–1605; < Latin spectrum; see spectrum] (http://www.thefreedictionary.com)

Sound spatialization offers the ability to composers and performers to specify how sounds are positioned in the listener’s audio field. Most consumer-quality audio systems allow for stereo fields (i.e., the ability to pan sound from left to right channel), becoming a standard household item in the 1970s. Around the same time, quadraphonic systems were introduced (i.e., making use of 4 channels), but did not meet with much commercial success, due to the industry’s indecision to create a standard. Within the 1980s and 1990s, surround sound was introduced into the consumer market with Cinema 5.1 (6 channels), and 7.1 (8 channels), e.g., DTS, Dolby Digital, etc. Still, these systems have proprietary formats, they require specialized software and hardware, and are not as readily available (as stereo systems). Additionally, various research techniques for sound spatialization were developed during this time, including Ambisonics and Vector Based Amplitude Panning (VBAP), which have not yet received commercial acceptance.

We present Specter, an open-source, scalable, easy to customize and deploy sound spatialization system, developed to facilitate music composition and live performance.

Specter was initially developed in the context of Time Jitters, a four-projector interactive installation (see Figure 1), designed by Los-Angeles-based visual artist Jody Zellen for the Halsey Institute of Contemporary Art in Charleston, SC, USA.1 Time Jitters includes two walls displaying looping video animation, and two walls with...
interactive elements. The concept is to create an immersive experience for participants, which confronts them with a bombardment of visual and sound elements. As participants enter the space, images and sounds are assigned to them. As participants move freely through the space, these images and sounds follow them. The result is an immersive, dynamic experience that unfolds in real-time as different people navigate the space. Several individuals contributed to this installation (including visual materials and overall concept design, interaction design, and sound design). In this paper, we focus on the design and implementation of Specter. Other aspects of this installation are presented elsewhere (e.g., [1, 2]).

Specter was designed to offer composers and performers a simple, expandable, and versatile environment to experiment with sound spatialization for music composition and live performance. It combines music information retrieval (MIR) with minimal hardware through the Open Sound Control (OSC) protocol to produce a low-cost, easily configurable and transportable system. Through various interfaces, users may position sounds at arbitrary locations and trajectories on a three-dimensional plane.

The system incorporates JythonMusic, an environment for symbolic music processing, music information retrieval, and live audio manipulation. It also utilizes Iannix, a graphical open-source sequencer for digital art for real-time generation, manipulation, and storing of sound trajectory scores. Finally, it uses Glaser, a sound manipulation instrument to render the various sounds in space by quickly manipulating their attributes.

The rest of the paper is organized as follows: Section 2 describes related background in sound spatialization. Section 3 defines the Specter system architecture; this includes a description of JythonMusic, the underlying music programming environment, used to implement Specter; Iannix, a graphical open-source sequencer for digital art, utilized to represent Specter trajectory scores; and Glaser, a sound rendering instrument. Section 4 presents a case study utilizing an MIR approach to generate a musical composition involving sound spatialization, which includes both static (pre-composed) and dynamic (interactive) sound trajectories, rendered with Specter. Finally, section 5 provides concluding remarks.

2. BACKGROUND

Although sound spatialization is a very promising field for developing new music and related composition techniques, it is highly underutilized (versus, say, timbre composition) because of the difficulty in exploring possibilities and performing existing compositions.

While there is much development already in timbre technologies for both analog and digital timbre, spatial computer music is being held back because composers have limited access to performance techniques and spaces with installed multi-channel systems [3].

Additionally, the software tools for spatial composition are few, compared to those for symbolic music, and for timbre compositions. Moreover, there is not a standard high-level format for representing spatial compositions or storing spatial recordings. As a result, spatial compositions may lose integrity and content as they are transferred from one technology to another, in order to be performed, given that the available performance spaces for spatial music are few, quite expensive to set up and maintain, have different architectures, and support different formats [4].

Nevertheless, sound spatialization is a rich field with many decades of research and development. Johnson et al. [5] provide a thorough overview of the early history of the field, starting with Schafer and Henry, who in 1951 performed the first pre-composed electroacoustic piece of music with dynamic spatialization at performance time. This was accomplished through a special interface controlling gain of individual speakers on a tetrahedral speaker array. Other important examples include construction of expensive performance spaces, during the 1970s and 1980s, such as the GRM Acousmonium, IMEB’s Gmebaphone, and the Univ. of Birmingham’s BEAST. Additionally, various spatialization algorithms have been developed for creating dynamic trajectories, and for spatial rendering for diffusion performance. They can be classified into two general categories:

(a) room-based diffusion, which involves programming autonomous spatial trajectories and complex spatial distribution patterns, through large numbers of speakers; one popular approach is Higher Order Ambisonics, e.g., see [6], and
(b) phantom-source positioning, which places less emphasis on amount of speakers, and focuses more on improving accuracy of sound object placement in the sound field, providing more control of sound trajectory rendering through better algorithms and data structures, and provision of interactive techniques for improved dynamic control of sound trajectories at performance time; one popular approach is Vector Based Amplitude Panning (VBAP), e.g. see [7].

Our system’s high-level architecture supports both approaches.

Lopez-Lezcano [8] discusses development of open, general-purpose sound diffusion systems. He identifies several important characteristics of such systems, in order for them to be more usable than the current state-of-the-art. These characteristics include simplicity, transparency, versatility, using commodity hardware, using free software, and having a small footprint. Our system is designed with these characteristics in mind, as described in the next section.

A significant research trend in sound spatialization involves gestural control (e.g., [9-12]) at composition time.
(such as through algorithms exploring dynamical systems, e.g., swarms and boids), but also at performance time (through specialized interfaces, such as data gloves, Kinect, and LeapMotion sensors, among others). Specter, through the underlying JythonMusic environment provides similar capabilities (a) through development of arbitrary algorithms to drive (or guide aspects of) music composition, and (b) through a variety of devices that can communicate with it via MIDI or OSC protocols.

3. SPECTER ARCHITECTURE

Specter incorporates three major components, JythonMusic, Iannix, and Glaser – all communicating via OSC to pass data and synchronize/coordinate their actions. The following sections describe each of the subsystems, and how they are combined to provide an environment to experiment with sound spatialization for music composition and live performance.

Through various interfaces and sensors, users may position sounds at arbitrary locations and trajectories on a two- or three-dimensional plane.

3.1 Music Information Retrieval

MIR functionality is available to Specter through JythonMusic, an environment for music analysis, composition and performance (see http://jythonmusic.org).

JythonMusic provides libraries for music making, image manipulation, building graphical user interfaces (GUIs), and for connecting computers to external MIDI and OSC devices, such as digital pianos, smartphones, and tablets.

JythonMusic is an outcome of a decade-long project exploring various aspects of music information retrieval, including investigation of fractals in music, and their relationship to human aesthetics (e.g., [13]). This on-going project explores Zipf’s Law (and related power laws) in music data mining, in music recommendation, and in music analysis, composition, and performance [14-16].

JythonMusic incorporates the following libraries. Primitives from each of these libraries are used in conjunction with Specter (as explained below) to create sound spatialization trajectories and related processes:

- **Music library** - provides primitives for creating music notes, phrases, parts, and scores, and for playing them live, as well as reading and writing them as MIDI or XML files.
- **Audio library** - provides primitives for loading and looping audio files, and for recording and looping live audio.
- **Zipf library** - provides primitives for extracting measurements from musical data (e.g., [13]).
- **MIDI library** - provides primitives for loading and looping MIDI files, and for connecting to external MIDI devices (e.g., pianos, guitars, synthesizers, etc.).

- **OSC library** - provides primitives for connecting to other devices via Open Sound Control (e.g., smartphones, tablets, computers, synthesizers, etc.).

Additionally, JythonMusic provides libraries for graphical interactivity, image manipulation and sonification, and event scheduling. Finally, it encapsulates various cross-platform libraries for MIR and music/sound manipulation, such as jMusic and jSyn.

In summary, JythonMusic provides the glue code through which Specter is implemented, and, through its libraries facilitates arbitrary possibilities for data mapping, sonification, and interaction.

3.2 Sound Spatialization Trajectories

Sound spatialization trajectories in Specter are modeled via Iannix, an open-source, 3D graphical sequencer for real-time generation, manipulation, and storing of musical and other scores (see http://www.iannix.org). As shown in Figure 2, Iannix scores consist of:

- **Curves**, which define spatial trajectories. These trajectories support cursors and triggers. Curves can be circular, straight lines, Bézier curves, free-form curves (drawn via mouse), or prescribed through math equations.
- **Cursors**, which follow the trajectories defined by curves, moving at a constant speed, when the score is played. Cursors report (via MIDI or OSC messages) their current coordinates in XYZ space (as defined by the Iannix score); also they can be set externally (via MIDI or OSC messages).
- **Triggers**, which report (again, via MIDI or OSC) when a cursor crosses them.

Additionally, through JythonMusic arbitrary math functions and algorithms may be implemented (such as boids, swarms, and other dynamic particle systems), which may
provide Specter with 3D trajectory information. This information may be stored as an Iannix score, or used in real-time to render sound spatialization trajectories.

3.3 Audio Rendering

Audio rendering in Specter is done through the Glaser subsystem. Glaser is an audio rendering instrument implemented using JythonMusic. It was developed for the TimeJitters exhibit (see section 1), and has been adapted to implement the functionality needed by Specter.

Glaser, in its basic form, allows exploring various sounds for sound design by manipulating their attributes (frequency, volume, and spatialization). It consists of three GUI displays with several sliders each, one per audio file. Through these, it allows a sound designer, composer, or performer to interactively control volume, frequency, and spatialization of an arbitrary number of audio files simultaneously.

Within Specter, the Glaser architecture has been extended to support both the Ambisonics and Vector Based Amplitude Panning approaches (e.g., see [6, 7]). This allows taking into account the spatial configuration (number of individual channels available) and geometry of the space, through existing algorithms, such as the ones used in [17].

3.4 Music Representation

Specter, through JythonMusic, utilizes a common-practice-based notation to represent audio to be rendered. This notation consists of:

• **Notes**, which specify pitch, duration, dynamic, and panning. For stereo, panning ranges from 0.0 to 1.0 (where 0.0 is left, 0.5 is center, and 1.0 is right of stereo field). Panning values greater than 1.0 are treated as identifiers for Iannix trajectories, which are used when the corresponding note is rendered. Pitches are used for sound frequency shifting (sounds are assigned a default / reference pitch, i.e., A4), durations are in seconds, and volume ranges from 0 to 127, following the MIDI standard.

• **Phrases**, which serve as containers for sequences of notes.

Additional, higher-level containers include parts and scores (again, see http://jythonmusic.org).

3.5 Expandable Architecture

The design concept behind Specter allows using easily accessible, low-cost equipment to render multi-channel audio, in an expandable architecture. This is facilitated by computer programming to account for the modular architecture.

Through the use of audio aggregates and low-cost, 2-channel USB audio interfaces, such as the Behringer UCB222 (approx. $30, at the time of this writing), it is possible to assemble a wide variety of sound spatialization architectures (e.g., obviously stereo, quadrophonic, 9-channel, 16-channel, and so on). Theoretically, any number of speakers / channels is possible for arbitrary sound spatialization installations.

3.6 Interfaces for Sound Spatialization

Given the underlying functionality provided to Specter via JythonMusic, a wide variety of interfaces may be used (or developed) to generate and/or capture sound trajectories. These trajectories may be stored (in a Iannix score) for later use in audio rendering, or be used immediately for real-time sound placement (this is exemplified in the case study presented in section 4). Possible interfaces include:

• **Kinect motion sensor**: Utilizing Kinect sensors with JythonMusic has already been implemented in the context of Kuarto, a motion-based framework for developing interactive music installations [1].

• **LeapMotion**: We have also developed a LeapMotion interface to capture fine movement of both hands. This inexpensive sensor, and its versatile API, allow for a wide variety of interfaces and associated gestures to be developed for natural, intuitive control of sound spatialization.

• **Smartphone**: Using smartphone sensors, e.g., gyro and accelerometer readings, one may develop various programs to control aspects of musical performance. One such example is presented in the next section.

Various other possibilities exist, utilizing any type of sensor that supports MIDI and OSC protocols.

4. MIGRANT - A CASE STUDY

*Migrant* is a cyclic piece for piano and computer, originally composed for Undomesticated, a public-art installation by Vassiliki Falkehag at Moore Farms Botanical Gardens, in the context of ArtFields 2015, held in Lake City, SC, USA (http://www.artfieldssc.org).

It is used here as an example of combining music information retrieval techniques and sound spatialization for music composition and live performance.

Migrant is part of the ISMIR 2015 music program to be performed on Wednesday, October 28, 2015 at the Sala Unicaja de Conciertos Maria Cristina in Málaga, Spain.

4.1 Composition Techniques

In terms of composition, Migrant integrates data sonification, interactivity, and sound spatialization.

The data used in the piece comes from migrant worker statistics, including migration patterns, age, wages, family dependents, and other elements of the migrant life experience. This data was collected from 56,976 in-person interviews with hired crop farm workers. The interviews
were conducted in 545 US counties and 43 states during fiscal years 1989-2012.2

Each note in the piece represents a single person. Melody, harmony and dynamic are all driven by the data. Data from 120 people were randomly selected. A few notes were manually adjusted by the composer to reflect his own aesthetic and migrant experience. Figures 3a and b show photos used by the composer to provide aesthetic inspiration for composing the sonification scheme.

The composition makes use of the golden ratio to affect the piece’s harmonic density (e.g., see Figure 4). The sonification code was written in JythonMusic using ideas presented in [20].

A preview of the piece (one cycle, mixed for two speakers) is available here – http://goo.gl/iYOVmY .

The composition employs interactivity to control tempo and spatialization of notes, via a smartphone-based controller manipulated by one of the performers. This controller sends gyro readings via OSC to a JythonMusic program, similar to this – http://goo.gl/dsTWFM .

4.2 Performance Needs

In terms of performance, the piece requires one piano, one computer, a video projector, two (or more) speakers (as described above), and a smartphone.

The original composition envisioned eight pianos arranged in 45-degree increments, with the audience seated in the middle. For ISMIR 2015, in order to demonstrate Specter, a single piano and a computer with sound spatialization are used.

Ideally, four speakers in a square configuration (as seen in Figure 5) allow the audience to fully experience the “flying around” of notes. However, two speakers in a stereo configuration work also, albeit losing one of the sound spatialization dimensions; in this case, the outcome is similar to the preview of the piece above.

4.3 Performance Instructions

Migrant is a cyclic piece for piano and computer, using a smartphone-based OSC controller, during the performance, to send these notes “flying around” in the sound field.

Each cycle of the piece lasts 4 minutes and 52 seconds. It is meant to be played in a continuous loop - minimally two times. For the ISMIR 2015 performance, the piece will be cycled exactly twice, for a total duration of 9 minutes and 44 seconds.

During the first cycle, the pianist plays the notes in the score verbatim. The computer layers identical notes in a computer-enhanced timbre and diffuses them in a circular pattern (or left-to-right pattern, for 2 speakers), thus generating a “flying-around” of notes.

The computer also displays images (using precise, scripted timings), via the video projector (again, see the piece preview provided here – http://goo.gl/iYOVmY ).

During the second cycle, the computer plays the complete first cycle (i.e., both the notes originally played by the pianist, as well as the enhanced timbres spatialized in the sound field).

The pianist is instructed to improvise additional notes, guided by the score notes. The only constraint is that an A natural minor scale is used. No constrains are given in terms of note start times, durations, or harmony. The pi-

---

anist is encouraged to create a musical narrative (to the best of their musical abilities – a challenge!), which aesthetically complements the sonified “narratives” of the people / notes in the data. In essence, this provides an opportunity for the pianist to interweave his or her own experience, aesthetic, and improvisatory skills into the piece.

During this, the computer performer utilizes the smartphone-based OSC controller to affect timing and spatialization of the computer-generated notes. This way, he or she controls aspects of the musical expression of the combined performance, through the following gestures:

- **Ready Position**: Smartphone is held facing up, parallel with the floor.
- **Controlling Tempo**: The phone is used in a percussive gesture (moving downward) to play the next note. When the phone pitch (see Figure 6) crosses the neutral (parallel to the floor) position, the next note is played.
- **Controlling Volume**: Device shake corresponds with loudness of notes. The more intensely one shakes or vibrates the phone as notes are generated, the louder the notes are.
- **Controlling Spatialization**: The yaw of the phone corresponds to placement of notes on the periphery of the sound field. (System is calibrated before the performance so magnetic north corresponds with Specter’s virtual north, as far as sound placement is concerned).

The interactive aspects of Migrant allow both human performers to musically interact with each other, and together, to interact with the musical “narrative” generated from the data.

5. **CONCLUSION**

Sound spatialization / diffusion systems normally require expensive, specialized equipment, which is usually hard to transport. We presented Specter, a simple, yet versatile environment to experiment with sound spatialization for music composition and live performance. By combining readily available hardware and software, through a simple, customizable architecture, we offer an inexpensive alternative to existing sound spatialization systems.

Specter may be used by MIR practitioners, as well as music composers and artists to explore and experiment with sound spatialization / diffusion more easily. Additionally, this project may facilitate development of innovative art installations, as well as new gaming experiences. Through the underlying JythonMusic system, developers may connect various MIR techniques to music composition and sound spatialization, open the door for new sonification applications, and develop innovative, immersive interactive applications (e.g., [21]).

**ACKNOWLEDGEMENTS**

We would like to thank Blake Stevens, Yiorgos Vassilandonakis, David Johnson, and Chris Benson for contributing to the development of some of the ideas and concepts presented herein. The first author would like to thank Vassiliki Falkehag for providing the inspiration for composing Migrant. Also, Mark Sloan and the Halsey Institute, in Charleston, SC, USA provided funding and space for the first implementation of Specter. Funding for this project has been provided in part by NSF (DUE-1323605).
REFERENCES


