LONG PAPERS

SOUND LABYRINTH: EXPLORATION OF THE EMBODIED SUBLIME THROUGH AN IMMERSIVE AUDIO/VISUAL INSTALLATION

Mark Pedersen
University of Melbourne

ABSTRACT

As immersive audio/visual technology continues to mature and become commercialised, the creation of sophisticated interactive systems that previously required significant infrastructure and funding comes within reach of the solo artist. With the ready availability of motion tracking systems like the Microsoft Kinect, and the proliferation of software components for creating immersive media environments, the challenge of audio/visual installation work is more than ever focused on addressing deeper conceptual issues, rather than solving technical problems.

Through the use of both representational and abstract audio, immersive sound spatialisation, multi-channel video, and the incorporation of gesture-based interaction, SoundLabyrinth applies theories of gesture within electro-acoustic composition, and theories of movement analysis and embodied music cognition, to the examination of the boundaries between virtuality and embodiment, transcendence and immanence, as an exploration of the “sublime within the everyday”.

1. INTRODUCTION

Immersion and interaction are two key objectives propelling digital art, requiring bigger screens, head mounted displays, multi-screen projections, 3D visualisation systems, ever more realistic rendering systems, and multi-channel surround sound systems that place the listener in the scene. Likewise the quest for interactive media has matured from keyboard-driven text adventures, to body sensing motion capture systems, including high fidelity systems such as Vicon-8 [10], through to domestic gesture tracking systems like the Microsoft Kinect.

While a thorough examination of the use of gestural interfaces within immersive installation practice is beyond the scope of this paper, it is worthwhile briefly touching on key developments and concepts. SoundLabyrinth draws upon the concepts and techniques of the now common immersive CAVE system [8], although with greater emphasis on immersive sound than fully immersive visuals, as described in Section 3. Such immersive environments are often experienced as trancelike, meditative, or mentally absorbing [11, p. 199], and as such, are more suited to the goals of SoundLabyrinth than a more open gallery architecture.

The other key development utilised by SoundLabyrinth is the natural body interface. By removing the need for any form of physical control apparatus, the distance between the participant and the virtual world of the artwork is reduced, heightening both the immersive quality of the work, and the sense of the participant’s embodiment within the work [ibid]. As a work exploring the interface between the embodied and the virtual, this distance reducing, boundary blurring technology is of great conceptual importance.

Rapid increases in available computing power, and the ubiquity and variety of user interface systems has reduced the cost of developing immersive environments. This ready availability increasingly enables work in this medium to explore conceptual issues, rather than focus on overcoming technical problems.

In the following sections, I first explore the conceptual and aesthetic issues which SoundLabyrinth seeks to address, before turning to a description of the work itself, the specific design strategies employed, and finally reflections upon the end result.

2. SUBLIME SOUND: EXPLORING THE SPECTRUM OF REPRESENTATIONAL AND ABSTRACT AUDIO

2.1. The sublime in (sound) art

This work arises from my desire to use sound as an artistic medium for exploring the sacred, not in an overtly religious sense, but in the sense of the (transcendent) sublime: “that which is beyond the senses”. Although having roots in antiquity, the concept of the sublime as an aesthetic polarity in contrast to beauty was most directly expounded by Edmund Burke[5]. Kant takes up this theme in A Critique of Judgement [12], noting that beauty “is connected with the form of the object”, having “boundaries”, while the sublime “is to be found in a formless object”, typified by “boundlessness”.

Schopenhauer [19] further developed Kant’s concepts of beauty and the sublime, in which the sublime lies beyond the subject’s ability to either physically cope with or mentally perceive or even
imagine. In grappling with this inability, the subject comes to apprehend the concept of incomprehensibility, and in this way becomes consciously detached from their own will.

What then is the relationship between the sacred and the sublime? Sacred, in its literal meaning, designates something set apart from common use, operating in a manner of its own distinctivity [17]. However the broader implication of the term is that it designates objects, places or actions which stand in relation to something beyond direct expression. If, after Schopenhauer, we consider the spectrum of aesthetic experience ranging from the sensate to the sublime, where the experience of sensate beauty relies upon perfecting representation, in contrast to the sublime, which is beyond representation, then sacred music or sacred sound is that which is not necessarily beautiful (not focused on the perfection of form), nor that which perfectly represents some externality, nor even that which perfectly represents itself, but that which engages the listener with the paradox of perceiving the unpersuadable, the unrepresentable and in that way offers detachment from self. Sacred sound, rather than being the concrete signifier of particular beliefs, is the liminal, aural space of Cobussen [6], within which there is the potential to experience the sublime.

From another perspective, the objective of seeking sacred sound, specifically in the context of digital sound making, is a response to Walter Benjamin’s critique of all art making conducted under the condition of mechanical reproduction [4]. Benjamin speaks of art being purely valued for its ritual function, often tied to a specific (religious) location, to being valued for its economic function as a globalised, reproducible commodity. In a time when the reproducibility and commoditisation of music has reduced its economic and cultural value to that of the size of one’s MP3 collection (rather than it’s musical or emotional appeal), it is perhaps more rewarding as a producer to offer a concrete representation of the Other, in as much as the unwanted signal calls for an openness, a reorientation of desire, a relinquishment of control, a reception, an entering into relationship with its source [13]. The next door neighbour’s party, and the prophet’s revelatory condemnation of societal injustice: both impinge upon a subjective contentment with the silence surrounding the status quo and offer an opportunity to engage - join the party, repent.

Sound art, as a field of expression that actively engages with the virtual (the abstracted gesture), and disrupts silence without necessarily offering a concrete representation of the Other, is a medium like no other in which to explore the transcendental sublime.

It is from this understanding of the aesthetics of the sublime that I approach the issue of sacred sound design: a practice of sound design that acknowledges a transcendental sublime but rather than attempting to either directly represent it, or express it through absence, engages in a continual questioning of the current moment, prompting awareness of the formlessness beyond perception, which is the ground of being.

2.2. Gesture and embodiment

Given the objective of exploring the sublime through sound art, to what extent can sound design transcend cultural context? The dialectic of essence and instance, virtual and actual, potential and event, becoming and being, finds its nexus in the embodied consciousness. Not the mind-body dualism of Descartes, but more so the body-mind continuum of Marc Leman [16, p. 82] suggests that it is the inclusion of “transcendence become immanent” in an work which lends uniqueness and interest to Benjamin’s terms, what is added is not just interest, but “aura”.

How then is sound a medium through which to explore the sublime Other? To what extent can sound be used to explore the generative and virtual soundscape of the virtual space of the sublime that Massumi alludes to?

That sound art is an ideal medium for exploring the theme of meaning which emerges from apparent chaos, or continual flux, is taken up by Cox [7]. Cox examines Liebniz’ concept of noise which gives rise to the idea that just as our movements arise from intentions (simulation of the movement), so perceptions of the external world map back to intentions because the trace left by the shared cognitive processes. This action-oriented ontology suggests that even at the social level, the actions of others are understood in terms our own intentions, i.e. our own perceptual activity. Thus, because individuals develop their own action-oriented ontology in a similar way by virtue of a common physiology, if not common culture, semantic communication is possible through music [14, p. 92].

This perspective is significant in the context of the discussion so far, for it provides both empirical evidence for and an explanatory theory which allows for a basic level of “objective” meaning to inhere in sound due to our common cognitive strategies, while still allowing for individual and cultural differences. In this sense, an embodied cognition approach to music making provides some resolution of the debate between formalists/structuralists and phenomenologists/post-structuralists.

Sacred sound, in its non-metaphysical formal forms of sound art to the morphology of human movement [21], an insight which is backed by work of Leman. From a compositional perspective, Smalley identifies several levels of engagement, prompting the listener to discern the source and cause through the quality of the material, the gestural archetype:

- **primal gesture**: basic proprioceptive gestural awareness, not linked to music making
- **first order**: recognisable sonic material subject to recognisable gestural play without levelling out the material
- **second order**: traditional instrumental musical performance
- **third order**: where a gesture is inferred or imagined in the music without the source (both the source and the material and the specific gesture are uncertain)
- **remote**: where “source and cause become unknown and unknowable as any human action becomes an action of nature”, but “some vestiges of gesture might still remain”, revealed by “those characteristics of effort and resistance perceived in the trajectory of gesture”.

In this regard, my approach to exploring the sublime through sound plays with both the spectrum of sonic material, from the representational to the abstract, and the spectrum of gesture, from the primal to the remote. In this way the spectrum of immemence-transcendence is likewise explored.

In the next section, I examine SoundLabyrinth, an immersive installation in which the dimensions of sublime sound are explored.

3. SOUNDLABYRINTH

SoundLabyrinth is focused on the relationship between sound and the body and the sense of meaning or “sacredness” that emerges from that relationship. The centrepiece of the project is an installation which uses ambisonic sound and immersive video projection set within large geodesic dome. As an installation, participants are able to explore a number of different sound worlds located within the space of the dome. These sound worlds comprise field recordings and sound designs by the author and material contributed by three poets: Nazid Kimmie (Australia/South Africa), Melike Ugezer (Australia/Turkey) and Rebecca Lemarie (India/Belgium).

The audio elements focus on sounds which relate, in broad terms, to the concepts of the sacred and the sublime, as well spoken word material relating to these themes.

The interactive mechanism of the installation allows participants to explore this sonic material as they move around within the dome, with the quality of the sound responding to their posture and gestures. Video projection within the dome provides additional context for the sound/gesture experience.

3.1. Audio/visual Infrastructure

The physical structure of the installation comprises a 6.5m diameter, three-frequency geodesic dome frame with a translucent white material cover. Within this frame, a 24.2 channel sound system is installed, along with 4 channel video projection. The 24.2 channel system consists of 24 matched mid-range Grover Notting CR-1 reference monitors [1] and 2 power subwoofers. The 24 mid-range speakers are laid out in three levels: 12 speakers in the bottom layer evenly distributed around the circumference of the dome (the ground), 10 speakers in the mid layer, approximately 2.5m from the ground, and 2 speakers directly overhead approximately 0.5m either side of the centre of the top of the dome. The crossover between the mid-ranges and subwoofers is at 100Hz, making the low frequency response used, and are distributed on either side of the dome so as to not disrupt the overall sound image. The signal structure for the system is shown in Figure 1. Figures 2 and 3 show the layout of components in plan and elevation respectively.

Spatialisation of the audio content is controlled via Max/MSP, using the ICST ambisonic externals [18]. Fourth order Farse-Malham encoding was found to give the best result in terms of sound image.
Imagine. In grappling with this inability, the subject comes to apprehend the concept of incomprehensibility, and in this way becomes consciously detached from their own will.

What then is the relationship between the sacred and the sublime? Sacred, in its literal meaning, designates something set apart from common use, especially in connection with divinity [17]. However the broader implication of the term is that it designates objects, places or actions which stand in relation to something beyond direct expression. If, after Schopenhauer, we consider the spectrum of aesthetic experience ranging from the senate to the sublime, where the experience of sensual beauty relies upon perfecting representation, in contrast to the sublime, which is beyond representation, then sacred music or sacred sound is that which is not necessarily beautiful (not focused on the perfection of form), nor that which perfectly represents some extremality, nor even that which perfectly represents itself, but that which engages the listener with the paradox of perceiving the unexpected, the unrepresentable, and in that way offers detachment from self. Sacred sound, rather than being the concrete signifier of particular beliefs, is the liminal, aural space of Cobussen [6], within which there is the potential to experience the sublime.

From another perspective, the objective of seeking the sacred through sound, specifically in the context of digital sound making, is a response to Walter Benjamin’s critique of all art making conducted in the context of digital reproduction [4]. Benjamin speaks of the transition from art being purely valued for its ritual function, often tied to a specific (religious) location, to being valued for its economic function as a globalised, reproducible commodity. In a time when the reproducibility and commoditisation of music has reduced its economic and cultural value to that of the size of one’s MP3 collection (rather than its aesthetic or cultural value), it is perhaps more rewarding as a producer to engage with the virtual (the abstracted, potentialised sound of the sublime, “beneath the threshold”).

Noise, in the more common sense of unwanted signal, also presents an opportunity for engagement with the Other, in as much as the unwanted signal calls for an openness, a reorientation of desire, a relinquishment of control, a reception, an entering into relationship with its source [13]. The next door neighbour’s party, and the prophet’s revelatory condemnation of societal injustice: both impinge upon a subjective contentment with the silence surrounding the status quo and offer an opportunity to engage - join the party, repent.

Sound art, as a field of expression that actively engages with the virtual (the abstracted, potentialised gesture), and disrupts silence without necessarily offering a concrete representation of the Other, is a medium like no other in which to explore the transcendent sublime.

It is from this understanding of the aesthetics of the sublime that I approach the issue of sacred sound design: a practice of sound design that acknowledges a transcendent sublime but rather than attempting to either represent it, or express it through absence, engages in a continual questioning of the present moment, prompting awareness of the formlessness beyond perception, which is the ground of being.

2.2. Gesture and embodiment

Given the objective of exploring the sublime through sound art, to what extent can sound design transcend cultural contexts? The dialectic of presence versus absence, virtual and actual, potential and event, becoming and being, finds its nexus in the embodied consciousness. Not the mind-body dualism of Descartes, but more of the body-mind continuum of Marc Leman.

For Leman, knowledge emerges out of need to act in the environment, not just collect information for its own sake. Herein lies the focus of the embodied approach is on action. Key to the embodied perspective is evidence for a close coupling between the cognitive processes for movement and perception. Leman [14, p. 37-102] provides extensive discussion of the evidence, including the behavioural observation of infants’ innate ability to perceive gestures and replicate them, and the neurobiological observation that some of the same neurons which are fired to create a gesture (e.g. grasping-with-the-hand) also fire when the subject observes another performing the same action (the so-called “mirror neurons”).

The tight coupling of movement and perception at a conceptual level gives rise to the idea that just as our movements arise from intentions (simulation of the movement), so perceptions of the external world map back to intentions because of the trace left by the shared cognitive processes. This action-oriented ontology suggests that even at the social level, the actions of others are understood in terms our own intentions, i.e. our own motivations are inferred from the movements of others. The presence of music are likewise attributed with intentionality because of the coupling of perception and movement. Thus, because individuals develop their own action-oriented ontology in a similar way by virtue of a common physiology, if not common culture, semantic communication is possible through music [14, p. 92].

This perspective is significant in the context of the discussion so far, for it provides both empirical evidence for and an explanatory theory which allows for a basic level of “objective” meaning to inhere in sound due to our common cognitive strategies, while still allowing for individual and cultural differences. In this sense, an embodied cognition approach to music meaning provides some resolution of the debate between formalist/structuralists and phenomenologists/post-structuralists.

This articulation of the spectromorphological forms of sound art to the morphology of human movement [21], an insight which is backed by work of Leman. From a compositional perspective, Smalley identifies several levels of gestural surrogacy - degrees of abstraction away from both the source material the gestural archetypes:

- **primal gesture**: basic proprioceptive gestural awareness, not linked to music making
- **first order**: recognisable sonic material subject to recognisable gestural play without intentional control
- **second order**: traditional instrumental musical performance
- **third order**: where a gesture is inferred or imagined in the music (both the source material and the specific gesture are uncertain).
- **remote**: where “source and cause become unknown and unknowable as any human action becomes ‘inconsequential’” but “some vestiges of gesture might still remain”, revealed by “those characteristics of effort and resistance perceived in the trajectory of gesture”.

In this regard, my approach to exploring the sublime through sound plays with both the spectrum of sonic material, from the representational to the abstract, and the spectrum of gesture, from the primal to the remote. In this way the spectrum of immiscible-transcendence is likewise explored.

In the next section, I examine Sound.Labyrinth as an immersive installation in which the dimensions of sublime sound are explored.
3.2. Audio content and spatialisation

The starting point for the audio content of the SoundLabyrinth lies in field recordings presented in a representational style, recognisably source-bonded. This material is used to construct a (hyper)realist soundscape: a distant creek bed, slightly to the right of the entrance, with bell birds in the middle distance, and wind chimes to the left. As participants journey through the physical space, the soundscape content shifts, allowing the visitor to pass through urban streetscapes, waterfront ports, desert zones, storms, cathedral interiors, and whatever spaces and locations may be represented by material contributed by the public. This material functions representationally and symbolically, evoking experiences of the natural and built environment, the sublime in the everyday.

Hidden within these sonic landscapes are fragments of poetic spoken word performances, multilingual expressions of the experience of engagement with the transcendent sublime, the prophetic voice of the Other. These spatialised voices are spatially elusive, remaining out of reach while the participant is in motion, but coming to rest in an immersive surround spatialisation if the participant stays still.

In addition to the representational field recordings and narrative/poetic spoken word, synthetically generated sound in the form of spectrally shifting drones, is used to mark entry into a zone of abstracted audio content, in which the participant’s gestures and posture interactively effect the soundscape. Gestural interaction is two-fold:
- at a passive level, the participant’s trajectory through the representational soundscape generates a spectrally sorted granular buffer constructed from the sequential concatenation of triggered field recordings;
- at an active, intentional level, the participant can scrub through this granular buffer.

The significance of the gestural control aspects of the installation are discussed further in Section 4, however before discussing these aspects, it is useful to cover the basic gestural sensor infrastructure.

3.3. Sensor System

SoundLabyrinth uses two Kinect depth sensing cameras: one installed overhead in the centre of the dome, and one installed horizontally at far side of the entrance (see Figure 2). The overhead camera provides input into a Max/MSP patch which uses a c-jit based blob tracking algorithm to provide participant locations as XY coordinates via Open Sound Control (OSC) to the main sound management and spatialisation patch. Figure 4 provides a more detailed view of the analysis and spatialisation subsystems. The depth-sensing feature of the Kinect camera simplifies the blob-tracking algorithm, as the camera subsystem can be set to ignore data below a threshold depth, eliminating the need to separate the subject from the background.

Position within the dome is used to smoothly transition between different soundscapes. The Max/MSP nodes object provides a zone-based trigger system which can also provided weighted output from each proximal trigger zone, allowing interpolation between triggered soundscapes.

3.4. Visual Elements

The visual elements of SoundLabyrinth consist of three display zones: one immediately above the horizontal Kinect sensor, as described above, and another two on either side of the entrance (see Figure 2). A fourth projector is used to highlight the structural elements of the dome itself, using full dome projection from a spherical mirror installed close to the entrance.

The visual elements of the installation are designed to heighten the participant’s sense of immersion in an alternate environment. The two lateral display zones act to visually evoke the sound worlds being explored, expanding the space contained within the physical dome to also include (the memory of) distant locations, while the distal display zone looks into an abstract realm, inviting exploration, and the full dome projection superimposes an ephemeral “virtual” dome structure on top of the physical structure, hinting at the enfolding of the potential into the actual.

Each display zone is responsive to the position of the participant within the space, changing content as the virtual soundscape change, providing a multi-sensory clue to the participant’s agency within the installation. Visual content for the display zones is handled by a separate system running VDMX[2] for content management and playback and Mad Mapper[3] for geometry adjustment. OSC data is sent from the main sound control patch to VDMX in order to trigger changes in content for each display zone.

The content of each display zone reflects the audio content of the soundscape, sometimes directly, such as images of water and reflection used in conjunction with the creek soundscape, while other scenes use more abstract visual impressions, such as layered footage of streetscapes and abstracted images of night time traffic used with urban ambience.
3.2. Audio content and spatialisation

The starting point for the audio content of the SoundLabyrinth lies in field recordings presented in a representational style, recognisably source-bonded. This material is used to construct a (hyper)realist soundscape: a distant creek bed, slightly to the right of the entrance, with bell birds in the middle distance, and wind chimes to the left. As participants journey through the physical space, the soundscape content shifts, allowing the visitor to pass through urban streetscapes, waterfront ports, desert zones, storms, cathedral interiors, and whatever spaces and locations may be represented by material contributed by the public. This material functions representationally and symbolically, evoking experiences of the natural and built environment, the sublime in the everyday.

Hidden within these sonic landscapes are fragments of poetic spoken word performances, multi-lingual expressions of the experience of engagement with the transcendent sublime, the prophetic voice of the Other. These spatialised voices are spatially elusive, moving in opposition to the participant, always remaining out of reach while the participant is in motion, but coming to rest in an immersive surround spatialisation if the participant stays still.

In addition to the representational field recordings and narrative/poetic spoken word, synthetically generated sound in the form of spectrally shifting drones, is used to mark entry into a zone of abstracted audio content, in which the participant’s gestures and posture interactively effect the soundscape. Gestural interaction is two-fold:

- A passive level, the participant’s trajectory through the representational soundscape generates a spectrally sorted granular buffer constructed from the sequential concatenation of triggered field recordings.
- An active, intentional level, the participant can scrub through this granular buffer.

The significance of the gestural control aspects of the installation are discussed further in Section 4, however before discussing these aspects, it is useful to cover the basic gestural sensor infrastructure.

3.3. Sensor System

SoundLabyrinth uses two Kinect depth sensing cameras: one installed overhead in the centre of the dome, and one installed horizontally at far side of the entrance (see Figure 2). The overhead camera provides input into a Max/MSP patch which uses a C++ based blob tracking algorithm to provide participant locations as XY coordinates via Open Sound Control (OSC) to the main sound management and spatialisation patch. Figure 4 provides a more detailed view of the analysis and spatialisation subsystems. The depth-sensing feature of the Kinect camera simplifies the blob-tracking algorithm, as the camera subsystem can be set to ignore data below a threshold depth, eliminating the need to separate the subject from the background.

Position within the dome is used to smoothly transition between different soundscapes. The Max/MSP nodes object provides a zone-based trigger system which can also provided weighted output from each proximal trigger zone, allowing interpolation between triggered soundscapes.

This approach provides a natural mechanism for exploration of the SoundLabyrinth, as triggered material fades in and out of hearing in response to position, alerting the participant to some level of agency.

3.4. Visual Elements

The visual elements of SoundLabyrinth consist of three display zones: one immediately above the horizontal Kinect sensor, as described above, and another two on either side of the entrance (see Figure 2). A fourth projector is used to highlight the structural elements of the dome itself, using full dome projection from a spherical mirror installed close to the entrance. The visual elements of the installation are designed to heighten the participant’s sense of immersion in an alternate environment. The two lateral display zones act to visually evoke the sound worlds being explored, expanding the space contained within the physical dome to also include (the memory of) distant locations, while the distal display zone looks into an abstract realm, inviting exploration, and the full dome projection superimposes an ephemeral “virtual” dome structure on top of the physical structure, hinting at the enfolding of the potential into the actual.

Each display zone is responsive to the position of the participant within the space, changing content as the virtual soundscape change, providing a multi-sensory clue to the participant’s agency within the installation. Visual content for the display zones is handled by a separate system running VDMX [2] for content management and playback and Mad Mapper [3] for geometry adjustment. OSC data is sent from the main sound control patch to VDMX in order to trigger changes in content for each display zone.

The content of each display zone reflects the audio content of the soundscape, sometimes directly, such as images of water and reflection used in conjunction with the creek soundscape, while other scenes use more abstract visual impressions, such as layered footage of streetscapes and abstracted images of night time traffic used with urban ambience.
4. GESTURAL INTERACTION AND EMBODIED EXPLORATION OF THE SUBLIME

In taking up the concept of the embodied sublime outlined in Section 2, the intersection of the abstract/virtual and the embodied/actual is the space upon which Sound.labyrinth focuses. The compositional approach is one of gradual intensification of this intersection. Elements of the installation’s soundscape respond to simple movement within the space, gradually progressing from that of a natural environment, to a landscape which is more abstracted (desert winds, snatches of poetry, synthetic drones) as the participant delves deeper into the space of the dome. In the point of greatest sonic abstraction, the participant encounters a new mode of interaction: one that responds to gesture, rather than just location. Through linking physical gesture and sound abstracted from the participant’s own journey through a more concrete, representational soundscape, there is an intensification of the inherent connection between sound and movement posited by Leman. The participant’s gesture is imprinted upon the (latent) sonic material, bringing it from the abstract into the actual. In parallel, the participant is immersed in sound: abstracted, virtualised material is not just triggered by movement, but forced upon the participant, but “brought to life” by them, in a sense, emerging from their movement.

There is a form of cybernetic inter-subjectivity that emerges as the participant, sensing the connection between sound and gesture, intensifies their exploration of the sounds world available to them, while the installation, sensing the participant responds to their gestures, either directly, or in oblique reference. Given that responses to gesture, rather than just location.

4.1. Correlating moving sonic forms and moving physical forms

It is therefore critical to base the gesture-sound mapping upon an analysis of movement and an analysis of sound which work together to enable the kind of multi-level surrogacy described by Smalley. In fact, because of the immediacy of the relationship between gesture and sound, surrogacy is no longer the appropriate term, as the sound gestural surrogacy is not replacing the physical gesture, but is being overlaid in superposition.

This is not to say that, in Smalley’s terms, that the nature of the interaction is only at the level second order surrogacy - that of the musical instrument. The intention of the gesture analysis process is not to create an interactive system which acts as a hyper-instrument based upon the participant’s body, but to observe potential relations between physical movement and sonic forms. The intention is that the sound design for the interactive system is informed by the body’s relationship to sound, but not to necessarily represent that relationship directly.

Rather, in Sound.labyrinth, the observed relationship between sound and body, both qualitative and quantitative, are used to select sonic elements which bear strong correspondences to physical movements of various types. In this sense the interactive element of the system operates in an improviser/collaborator paradigm, rather than in a performer/instrument paradigm.

Drawing upon Laban Movement Analysis (LMA) [22] I sought to identify correlations between Laban Effort Shapes, as a canonical set of movement qualities, and the spectromorphological attributes of sound developed by Smalley[21]. Marshall [15] takes a similar approach in the application of Smalley’s spectromorphology to the design of an interactive sensor-based sonification system.

In other work comparing Laban and Smalley [9], we observe that it is possible to derive some correspondences between the two models: time in LMA relates to both the motion launching (or attack) and the texture quality of spectromorphology; weight relates to the rootedness quality; and space relates to contour.

This mapping is imperfect, but is sufficient for us to design the sensing system of Sound.labyrinth to support a relatively intuitive level of engagement with the abstracted sound material through gesture.

Expressing Smalley’s spectromorphological movement qualities in terms of more common synthesis parameters, we have taken the following approach to selecting synthesis parameters which correspond to various LMA movement qualities. These are summarized in Table 1.

5. REFLECTIONS

In reflecting upon Sound.labyrinth as a work oriented toward enabling exploration of, or encounter with the sublime, a number of layers to the work emerge as significant.

If the representational sound content operates at the level of memory and association (recalled experience), and the poetic spoken word operates at a linguistic level of thought (abstracted experience), the synthetic, gestural content operates at the level of sensation (immediate experience), that is, at the level of the sub-liminal. In Smalley’s terms, this material operates at the level of first order gestural surrogacy. The representational sound is at once specific (concrete), as a recording of an actual instance of water flowing, birds calling, etc., and general (abstract), as a type of any such instance, by virtue of the generalising operation of human memory extending the specific instance to the abstract.

The same material, granularised, is at once abstract, in as much as it has been disconnected from its source and is “unrecognisable”, and concrete, in the sense of being only sound, disconnected from the objectifying process of recognition.

Upon this material, the specific gestural patterns of the participant, in the form of their exploratory trajectory (their specific personal history) as well as their immediate physical gestures, are imprinted. This forms the subliminal layer of sound content, operating beneath the threshold of thought, beneath the threshold of recognition, at the third or, even remote, level of gestural surrogacy. The participant, as an immersed subject within the immanent field of the installation, embodies the sublime.

<table>
<thead>
<tr>
<th>Motion Factor</th>
<th>Movement Quality</th>
<th>Synthesis properties</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>Quick</td>
<td>short percussive sounds(short onset, short continuation)</td>
</tr>
<tr>
<td></td>
<td>Sustained</td>
<td>sustained sounds(longer onset, long continuation, long decay)</td>
</tr>
<tr>
<td>Weight</td>
<td>Light</td>
<td>low density sounds with harmonic content and greater reverb, higher pitch</td>
</tr>
<tr>
<td></td>
<td>Strong</td>
<td>high density sounds with greater noise, less reverb, lower pitch</td>
</tr>
<tr>
<td>Space</td>
<td>Direct</td>
<td>steady pitch and spectra, less reverb</td>
</tr>
<tr>
<td></td>
<td>Flexible</td>
<td>varying pitch and spectra, more reverb</td>
</tr>
</tbody>
</table>

Table 1. LMA Motion Factors and sound synthesis properties

This mapping is used bi-directionally within Sound.labyrinth: for participant gesture to sound, as a responsive recognition by the system of participant’s embodied expression; and from the system toward the participant, as a prompt toward alternative gestures, postures and positions.

Using spectral analysis and granular, concatenative synthesis [20], the original field recordings triggered by the participant during their journey through Sound.labyrinth are dissected into component particles, blended with synthetic sounds, and shaped in response to gestural input.

The gesture-sound mapping described in Table 1 is used to modify the granulation parameters of both the concatenative synthesis engine and a hybrid subtractive synthesis / sample manipulation engine in order to produce the desired sound quality (short, percussive; sustained; light; dense). Spectral selection of grains is also influenced by direct position, where physically higher gestures produced higher pitched material, and vice versa. The use of a mixture of gesture and space mapping as well as effort quality analysis provides both immediacy, which helps the participant recognise their agency with respect to sound generation, as well as subtlety and depth, allowing for deeper exploration of the sound using different movement qualities.

The temporal qualities of the participant’s journey through the physical space of the Soundlabyrinth are also preserved within the granular buffer of the concatenative synthesis engine, allowing the various sound worlds encountered to be explored at a new level of perspective - from within the sound material - while remaining vestigially recognisable through the temporal grouping of source-related material.
In taking up the concept of the embodied sublime outlined in Section 2, the intersection of the abstract/virtual and the embodied/actual is the space upon which SoundLabyrinth focuses. The compositional approach is one of gradual intensification of this intersection. Elements of the installation’s soundscape respond to simple movement within the space, gradually progressing from that of a natural environment, to a landscape which is more abstracted (desert winds, snatches of poetry, synthetic drones) as the participant delves deeper into the space of the dome.

At the point of greatest sonic abstraction, the participant encounters a new mode of interaction: one that responds to gesture, rather than just location. Through linking physical gesture and sound abstracted from the participant’s own journey through a more concrete, representational soundscape, there is an intensification of the inherent connection between sound and movement posited by Leman. The participant’s gesture is imprinted upon the (latent) sonic material, bringing it from the abstract into the actual. In parallel, the participant is immersed in sound: abstracted, virtualised material is not just triggered by movement, not forced upon the participant, but “brought to life” by them, in a sense, emerging from their movement.

There is a feedback of cybernetic inter-subjectivity that emerges as the participant, sensing the connection between sound and gesture, intensifies their exploration of the sound world available to them, while the installation, sensing the participant responds to their gestures, either directly, or in oblique reference. Given the casual nature of the relationship, this temporary fusing of participant and installation as a combined performative system would fail if the mapping between gesture and sound was not grounded in the embodied intelligence of the participant. Without an awareness of the relationship between gesture and sound, the sounds generated in response to gesture could feel alien and disconnected, and fail to invite further exploration.

4.1. Correlating moving sonic forms and moving physical forms

It is therefore critical to base the gesture-sound mapping upon an analysis of movement and an analysis of sound which work together to enable the kind of multi-level surrogacy described by Smalley. In fact, because of the immediacy of the relationship between gesture and sound, surrogacy is no longer the appropriate term, as the sound gesture is not replacing the physical gesture, but is being overlaid in superposition.

This is not to say that, in Smalley’s terms, that the nature of the interaction is only at the level second order surrogacy - that of the musical instrument. The intention of the gesture analysis process is not to create an interactive system which acts as a hyper-instrument based upon the participant’s body, but to observe potential relations between physical movement and sonic forms. The intention is that the sound design for the interactive system is informed by the body’s relationship to sound, but not to necessarily represent that relationship directly.

Rather, in SoundLabyrinth, the observed relationship between sound and body, both qualitative and quantitative, are used to select sonic elements which bear strong correlations to physical movements of various types. In this sense the interactive element of the system operates in a improviser/collaborator paradigm, rather than in a performer/instrument paradigm.

Drawing upon Laban Movement Analysis (LMA) [22] I sought to identify correlations between Laban Effort Shapes, as a canonical set of movement qualities, and the spectromorphological attributes of sound developed by Smalley[21]. Marshall [15] takes a similar approach in the application of Smalley’s spectromorphology to the design of an interactive sensor-based sonification system.

In other work comparing Laban and Smalley [9], we observe that it is possible to derive some connections between the two models: time in LMA relates to both the motion launching (or attack) and the texture quality of spectromorphology; weight relates to the rootness quality; and space relates to contour.

This mapping is imperfect, but is sufficient for us to design the sensing system of SoundLabyrinth to support a relatively intuitively level of engagement with the abstracted sound material through gesture. Expressing Smalley’s spectromorphological movement qualities in terms of more common synthesis parameters, we have taken the following approach to selecting synthesis parameters which correspond to various LMA movement qualities. These are summarized in Table 1.

<table>
<thead>
<tr>
<th>Motion Factor</th>
<th>Movement Quality</th>
<th>Synthesis properties</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>Quick</td>
<td>short percussive sounds (short onset, short continuation)</td>
</tr>
<tr>
<td></td>
<td>Sustained</td>
<td>sustained sounds (longer onset, long continuation, long decay)</td>
</tr>
<tr>
<td>Weight</td>
<td>Light</td>
<td>low density sounds with harmonic content and greater reverb, higher pitch</td>
</tr>
<tr>
<td></td>
<td>Strong</td>
<td>high density sounds with greater noise, less reverb, lower pitch</td>
</tr>
<tr>
<td>Space</td>
<td>Direct</td>
<td>steady pitch and spectra, less reverb</td>
</tr>
<tr>
<td></td>
<td>Flexible</td>
<td>varying pitch and spectra, more reverb</td>
</tr>
</tbody>
</table>

Table 1. LMA Motion Factors and sound synthesis properties

This mapping is used bidirectionally within SoundLabyrinth: from participant gesture to sound, as a responsive recognition by the system of participant’s embodied expression; and from the system toward the participant, as a prompt toward alternative gestures, postures and positions.

Using spectral analysis and granular, concatenative synthesis [20], the original field recordings triggered by the participant during their journey through SoundLabyrinth are dissected into component particles, blended with synthetic sounds, and shaped in response to gestural input.

The gesture-sound mapping described in Table 1 is used to modify the granulation parameters of both the concatenated synthesize engine and a hybrid subtractive synthesis / sample manipulation engine in order to produce the desired sound quality (short, percussive; sustained; light; dense). Spectral selection of grains is also influenced by direct position, where physically higher gestures produced higher pitched material, and vice versa. The use of a mixture of gesture-sound mapping as well as effort quality analysis provides both immediacy, which helps the participant recognize their agency with respect to sound generation, as well as subtlety and depth, allowing for deeper exploration of the sound using different movement qualities.

The temporal qualities of the participant’s journey through the physical space of the SoundLabyrinth are also preserved within the granular buffer of the concatenated synthesis engine, allowing the various sound worlds encountered to be explored at a new level of perspective - from within the sound material - while remaining vestigially recognisable through the temporal grouping of source-related material.

5. REFLECTIONS

In reflecting upon SoundLabyrinth as a work oriented toward enabling exploration of, or encounter with the sublime, a number of layers to the work emerge as significant.

If the representational sound content operates at the level of memory and association (recalled experience), and the poetic spoken word operates at a linguistic level of thought (abstracted experience), the synthetic, gesturally encoded content operates at the level of sensation (immediate experience), that is, at the level of the sub-liminal. In Smalley’s terms, this material operates at the level of first order gestural surrogacy.

The representational sound is at once specific (concrete), as a recording of an actual instance of water flowing, birds calling, etcetera; and general (abstract), as a type of any such instance, by virtue of the generalising operation of human memory extending the specific instance to general categories.

The same material, granularised, is at once abstract, in as much as it has been disconnected from its source and is “unrecognisable”, and concrete, in the sense of being only sound, disconnected from the objectifying process of recognition.

Upon this material, the specific gestural patterns of the participant, in the form of their exploratory trajectory (their specific personal history) as well as their immediate physical gestures, are imprinted. This forms the sublime layer of sound content, operating beneath the threshold of thought, beneath the threshold of recognition, at the third or, even remote, level of gestural surrogacy. The participant, as an immersed subject within the immanent field of the installation, embodies the sublime.
Participant reactions to the work were quite varied. Common feedback from participants included comments on the meditative quality of the installation, that they found it relaxing, disorienting or trippy. Some wanted to immediately sit still, lie down, and be passively immersed in the sound field, rather than actively exploring it. Others were initially overactive in their movement and took some time to discover the subtleties of the work that are revealed from more careful movement and exploration within the space. The immersive quality of the audio was frequently commented on.

Some participants remained unaware of the shift to the deeper gestural control mode, primarily because they continued to rapidly explore the space, and thus moved in and out of the gestural control zone before noticing the additional control layer. Others found the spatial zone added another layer of engagement and spent time exploring this aspect of the work.

Engagement with the visual elements of the installation was strong, always drawing comment from participants. A number of participants connected with the linguistic layers of the audio content, enquiring about the languages used and the content and origin of the texts.

The use of the geodesic dome structure itself also drew strong reactions, with some participants remarking that they experienced a unique shift in feeling as soon as they stepped inside the dome, and often connected the spatial qualities of the installation with feelings of coalescence, calmness and meditation.

From my own perspective, Soundlabynth functions as a powerful space in which to work with sound. The immersive qualities of three-dimensional surround sound connected to gesture and responsive to movement within the space open up the deeper conceptual space of the interface between embodiment and the abstract. The sublime, by its nature, can never be grasped, and there is still much more to explore.

6. REFERENCES


A TEMPORAL GENERATIVE GRAPH GRAMMAR FOR HARMONIC AND METRICAL STRUCTURE

Donya Quick, Paul Hudak
Yale University
Department of Computer Science
New Haven, CT USA
donya.quick@yale.edu, paul.hudak@yale.edu

ABSTRACT

Most grammars that have been proposed for automated music composition fall into conventional linguistic categories, such as context-free, context-sensitive, and probabilistic versions of each. For parsing (i.e. musical analysis) these distinctions are important, because of the computational complexity of parsing using these different grammars. But, for generation (i.e. derivation), the complexity issues are sometimes different, and other goals for having a generative grammar come into play.

In this paper we describe a new category of grammar for generating abstract harmonic and metrical structure. This class of grammars has two distinctive features. First, it is temporal, meaning that production rules are parameterized by the durations of phrases, thus allowing us to express the metrical structure of a composition. Second, it is a graph grammar, meaning that the parse trees (or derivations) are actually graphs allowing shared nodes, thus enabling us to express the sharing, i.e. repetition, of specific musical phrases.

We formally define this class of grammars, describe our generative implementation of it, and present a realistic example of its use: a specific grammar tailored to some styles of classical Western music. In addition, we show that this class of grammars integrates nicely with the notion of chord spaces to generate concrete chords from the abstract harmonic structure generated from the grammar.

1. INTRODUCTION

Automated composition is a complex task. Music is multi-dimensional, and the solution spaces are very large. Even a seemingly simple task such as choosing pitches for chords results in an exponential growth in computational complexity. Having elegant and efficient algorithms is therefore important for traversing these solution spaces. As a further complication, those large solution spaces contain many poor or undesirable solutions. So, not only does a generative algorithm for music have to be efficient, it has to also find satisfactory solutions in a sea of noise.

For example, probabilistic context-free grammars (PCFGs) are both efficient to use generatively and, if constructed appropriately, are capable of producing satisfactory solutions. But they also generate many unsatisfactory solutions. We could do better if at least some notions of ‘satisfactory’ could be captured in the grammar itself, thus eliminating large undesirable regions of the solution space.

To address this problem, we present a new category of grammar that we refer to as a temporal generative graph grammar (TGGG). TGGGs are powerful enough to express both temporal and sharing constraints, leading to an effective method for the automated generation of harmonic and metrical structures of music.

There are two distinctive features of a TGGG. First, one often wishes to preserve certain metrical constraints in a composition. For example, one may wish to replace a I chord with a IV I V progression, but with the constraint that the total duration of the IV I V progression is the same as that of the original I chord. In our TGGG framework, this is easily specified with the rule $I \rightarrow IV \cdot I \cdot V \cdot I$.

This allows metrical constraints such as those found in Generative Theory of Tonal Music (GTTM) [14] to be captured by the grammar itself in a more formal way.

Technically, this feature results in an infinite number of production rules, but, when used in a generative setting, that is not problematic. The initial duration $t_0$ associated with a given grammar’s start symbol directly determines all the others. One can think of each rule as being a function that takes a time (duration) as an argument.

The second distinctive feature of a TGGG is that it captures the notion of ‘repetition of a phrase’, i.e., the sharing of a particular sentential form in a derivation. For example, if part of a composition has the form $ABA$, and we expect both occurrences of $A$ to be precisely the same phrase, we can write the following in a TGGG (where we have omitted any temporal superscripts for simplicity):

$$x \rightarrow \text{let } y = x \text{ in } x A x$$

(1) where $S$, $A$, and $B$ are nonterminals. Note that ‘$A$’ in this case means that some arbitrary number of production rules are applied to the nonterminal $A$, and that the result is used identically in the two occurrences of $x$ (versus simply writing $ABA$, where each $A$ could expand differently).

Therefore, a derivation tree is actually a more general type of directed graph, where some of the nodes share edges. These shared nodes occur whenever there is a let expression.

Technically, this means that a TGGG is a context-