Sound object and space: developing concepts by making software

Bachelor's thesis

Julius Raskevicius

Institute of Sonology, 2016

Abstract

Concepts of sound object as a mental and programmed representation of sound are explored in the thesis through compositions and software. Terms “instance” and “class”, as seen in object-oriented programming, are used to describe a unit to compose and analyze acousmatic music. The role of spatiality and its relationship to vision is also discussed. Three computer programs are presented, each marking the progression in the development of sound object from spatial perspective.

Acknowledgments

I would like to thank all the teachers in the Institute of Sonology for both the knowledge and inspiration I received throughout the studies. I am also equally thankful to other students of sonology whose music and research plays an important role in my artistic direction.

Contents

1 Listening

1.1 Introduction

> "Aesthetics deal with perception, and thus more with the listener than the composer..." (Koenig, 1968)

Acousmatic music offered me a new experience: lacking their visual causes, sounds could be perceived in a more vibrant and imaginative way, creating mental images and shapes in the form of imagined sources. These abstract entities took volumetric form and were situated in a specific space which was not rigidly related to the setting in which the sound was diffused.

It seems that spatial experience is more related to the state of mind, than a particular genre or school of composition, as I experienced such moments in acoustic and electronic music of various periods. A musique concrète piece by Hanna Hartman Die Schrauben, die die Welt zusammenhalten (2002) provokes a very visual experience each time I listen to it. The piece starts with a scattered sequence of crackles and twists of the invisible machine – traversing from soft and brittle to vast and deep impact sounds, each delimiting the edges of the imagined space. The volume here is created by giving the listener spatial hints from differently distanced mics during the recording of acoustic sources, as well as differently sized impacts on a resonant wooden body. Rest-wise the piece also leaves enough negative space for detailed decaying tails to unravel. The sounds are exposed in discrete and transparent way, providing a lot of contrast to the sound groups which follow afterwards. An example of spatially-rich electronic music could be Kontakte (1960) by Karlheinz Stockhauzen. One aspect of the piece is its expressive use of sound projection, which creates space forms as the identifying aspects of the moment. Pieces from spectromorphological school also played an important role in raising questions about structure and building unit of acousmatic music. For example, a piece by Manuella Blackburn, Kitchen Alchemy (2007), is built up from events which are both material and spatially centered.

Even though a lot of the concepts discussed in this thesis are visual, most of the piece examples are to be listened without seeing the visual sources. Instead of using visual terms purely as metaphor, I would like to argue that the two modalities are tightly linked in the memory of the experience, to the degree it is possible to speak about sound as an observed physical object. The degree of the integration between audition and vision is probably a strongly individual thing, but discussing sound with visual and spatial terminology is still useful from practical standpoint. Metaphor of object has strengths in organization of material and opens compositional directions dealing with space. Integration of perceptual modes could be discussed through concepts of sound object and role of spatiality in the context of acousmatic listening.

1.2 Disambiguation of Sound Object

The perceptual experience of space discussed previously is related to the division of sounds into separate entities, which are referred to as sound objects in the thesis. The term "sound object" is ambiguous and distinction between three separate concepts can be made: Object of sound describes a physical object-like nature of an entity, emphasizing its discreteness, its volumetric occupancy of space, its ability to be interacted with physically (through human gesture or inbetween inanimate sources) and its ability to produce sound as a result of agental energy transfer; Sound object as a structural unit in composition, emphasizing structural divisibility and separation into exchangable behaviors. The term focuses on sound as part of a structure and its ability to connect with other parts. Objet sonore, as coined by Pierre Shaeffer, referring to morphologically distinct class of sounds, united by certain features. It refers to an idealized version of sound object.

A distinction between representation of a sonic experience and representation as a model can also be made. While listening to acousmatic music, we experience ambiguous sounds and represent them in memory as supposed object of source, either by involuntary association or as a method to organize abstract material. While composing an opposite direction is taken – a programmed model of experienced sound object is used to generate sounds. If we suppose that somewhere sounds exist objectively or purely sonically, a programmed sound object is a representation of a representation.

"Sound object" as used in the thesis is a hybrid of all three concepts. From the listener's perspective, sound object is an experience which is created when a group of sounds occur together and allude to material properties. The generality of the sounds, especially when presented as a sequence of variations, defines an identity of sound object. From the composer's perspective, sound object is defined by the structure which has adjacent key points in time (sounds start together) and in space (sounds start at the same location). Instead of being called a "sound group" or "sound class", the "object" side of the term implies association to a physical object. Our past experience of an energy transfer in material world – impacts, material deformation, caused movement – leads to expectation that concurrently sounding components will belong to the same object.

Integration between perceptual modalities has evolutionary roots. In natural environment organisms must solve the problem of navigation. Senses of vision, audition, proprioception, haptic feedback and others are used when attempting to traverse terrain. Transfer of energy in a material object happens in multiple directions simultaneously: changes in position and velocity are perceived visually as motion; change to the surrounding medium – air or liquid – are perceived as sound or haptic vibration. As a result, a change in the environment naturally produces multiple correlated streams of information: *"In a natural environment distinct sources of auditory and visual information are integrated seamlessly to form a unified representation of external space that guides navigation"* (Neuhoff, 2004). In perception, information from all the senses is usually experienced together, thus it could be expected that the opposite may also be true – information from all of the senses will be expected even when some senses are not part of a stimulus. As Alva Noe notes, *"To perceive you must be in posession of sensory-motor body skill."* (Noe, 2004), indicating the corporeal aspect of perception – a sound may have to be perceived through its supposed cause.

Corporeity of the sound object introduces the concept of space. As Smalley notes about the role of all senses in perceiving a sonic event, *"All these modes embody underlying spatial attributes: texture has space, gesture operates in spaces integrated into the gestural task, cultural and natural scenes are spatial, the highs and lows and motions of sound spectra evoke space"* (Smalley, 2007). Thus sound object is located in space as well as defined by space.

1.3 Roots of the concept – objet sonore

It is possible to track the development of the term "sound object" in Shaeffer's journals over time. The development is relevant to understand three components of sound object described above. At the beginning of his studies in concrete music sound object refers to a *"physical-material thing – a source for the production of sound"*; later, with registration of sound on disc, it gains more abstract form of *"something discrete and complete, the fruit of a mode of considering or listening to the fragment torn from the whole"*; lastly the term is developed into the most abstract form, in line with phenomenological ontology of Husserl, requiring specific conditions and mental practice to be revealed (Kane, 2014). A practice of eidetic reduction encourages thinker to produce a series of versions of the sound and in the process it *"becomes evident that the unity runs through multiplicity"* (Russell, 2006)

There is a rationale behind the development of sound object towards its more abstract form. As Brian Kane argues two recurrent desires of Pierre Shaeffer's experiments were *"a compositional desire to construct music from concrete objects <...>, and a theoretical desire to find a vocabulary, solfege, or method upon which to ground such music"* (Kane, in. cit.). The first one could be seen as an intent to departure from musical aesthetic of the time, a look for a new sound, and the second as a consolidation of the new aesthetic's intentionality. I see the second point as a desire for an unfamiliar sound to be perceived in the frame of its own aesthetic system, where its morphological attributes are the subject of focus, instead of them simply cuing in recognizable material sources. In the context of these motivations, I would like to discuss the practice of acousmatic listening.

Does acousmatic listening conflict with visualization of sound?

> "Acousmatic, adjective: a sound that one hears without seeing what causes it." (Shaeffer, translation by Cox and Warner, 2004)

Acousmatic listening intentionally removes visual cues as to obscure the cause of the sound. The relationship between sound object and acousmatic listening is described in the following statement by Shaeffer - *"The sound object is never revealed clearly except in the acousmatic experience"* (Kane, in. cit.) Shaeffer explains the need to avoid seeing the source as a way to focus on a specific mode of listening – entendre - which is summarized by Brian Kane as *"... listening to sound's morphological attributes without reference to its spatial location, source, or cause"* (Kane, ibid).

It is possible to explain the concerns of Shaeffer through cognitive science studies on perception and attention. Vision has been shown to interact with auditory and other modalities, dominating over sonic information at the stage of integrating information into unified representation. Given the strength of effect visual information has on perception in general, it is not surprising that image is capable of drawing attention away from the morphological attributes of sound, which forms the focus of interest in my work.

Vision has an overriding role to other modalities. It has been shown that vision dominates over kinaesthetic feedback in judging the straightness of surface (Gibson, 1933) In the experiment, test subjects had to wear image-bending glasses for a period of 1 hour. Typical effect of applying visual distortions for a long time is adaptation to the image – in previous experiments it has been shown that objects which initially appear vertically bent, are perceived as straight after the vision adapts. When the special glasses are removed, an after-effect is experienced, causing surrounding objects to appear bent in the opposite direction. Gibson has demonstrated that the effect of vision bending glasses is not reduced by touching the edge of a test surface – subjects consistently report that the edge feels actually bent. Thus no conflict is experienced in the perception of incongruent evidence between the modalities, in some cases certain perceptual information is simply ignored.

Dominating effects in perception of vocal utterances have also been observed. An appropriately named study Hearing lips and Seeing Voices (McGurk, McDonald, 1979) demonstrates how vision may distort recognition of auditory utterance. In the experiment subjects have been shown a video of a person talking. Visually, the movement of lips pronounced an utterance [ga], whereas the video was dubbed with a voice saying [ba]. The subjects later recalled hearing [da], demonstrating the interaction between the two domains.

Tendency of sound to yield to visual stimuli can be seen in everyday life. Michel Chion points out two cases: localization of a passing jet plane and speaker position in cinema. In the first example, when a mismatch occurs between sound of a plane passing above and location of its image, localization sometimes jumps from perceived location to the location of the image. In the second example, the effect is so strong that we no longer see it as anything strange – when a sound is projected in the cinema, its location never matches the location of the source, yet we identify it as coming from the lips of the speaker. Sound becomes spatially bound to the visual source (Chion and Gorbman, 1994).

Motivation to listen to music in acousmatic conditions can come from two factors. First, in perceptual integration visual cues provide enough evidence to clearly and unambiguously identify the sound source, thus acting as opposing evidence in all cases where sound is processed or otherwise presented in irregular way. In acousmatic music we often encounter situations which are beyond normal listening conditions. In Die Schrauben, die die Welt zusammenhalten hyper-real or surreal image is formed as sounds do not match the dimensions of their acoustic sources. Sound is over-emphasized to put the focus on its microstructure. In this case, presenting the acoustic source would act as incongruent evidence in creation of the perceived object representation, thus negating the microsonic effect. Alternatively, if sound was varied (mentally or through editing) with a compositional goal of exposing an identifying sonic characteristic (as in the process of revealing objet sonore), visual cues could counteract variation.

Second, if the source is identified clearly, the morphological complexity of the sound can be substituted with the symbolic representation of the object itself, so that the sonic properties are no longer relevant. Shaeffer has defined separate listening modes that are different in the intent of the listener. As noted by Kane, comprendre *"refers specifically to the reception of sounds mediated by sign systems or languages"*, which includes musical languages, while écouter describes *"mode of listening that is securely bound to the natural attitude, where sounds are heard immediately as indices of objects and events in the world"*. Both of the modes involve signification and thus "bypass" the need to be aware of the sonic complexity of a sound.

The definition of entendre separates sonic characteristics from other identifying aspects - *"spatial location, source or cause"*. The exclusion of "location" as a parameter of sound object strikes as curious. It could be understood in case of acoustic sounds, especially when non-traditional sources are used – it may be useful to remove location for abstraction and to avoid identification. However, in modern context the effect of perceived location of sound is not necessarily revealing its cause, simply because synthesized sounds do not have any material causes. Space is only necessarily conceptualized as separate from the source object in acoustic situation alone. Once the sound is completely synthetic the space and the sound can be inherently connected and sounds can be classified by their manifestation in different trajectories. For example, a sound, which moves within certain dimensions of space can have a unique identity from a similar sound which is spatially static – the *"unity through multiplicity"* would be of spatial type, connecting sounds which have certain position, size or trajectory.

In my experience, interpreting abstract material as it would be bound to a physical cause acts as an organizing factor and aids comprehension of music and appreciation of sound. Integration of senses

Figure 1: DeepDream algorithm creating visual pareidolia (Wikimedia, 2016)

has a positive effect in creating representation of sound object.

1.4 Adding visuals to sound

If technological advances in sound synthesis lets us create sounds without actual material sources, then perhaps the acousmatic listening is no longer a necessary condition to experience morphological intricacy of sound? I believe there are other reasons why image could be avoided – one of the strengths of sound as an artistic medium is the ability to create a strong impact which is nevertheless objectively fleeting and prone to ambiguity. The play on the identity of sound sources, even with synthetic material, is a common subject of acousmatic composition. It could be said that perceiving sound object is a form of pareidolia – a phenomenon when mind perceives a familiar pattern in something that does not contain it (Pareidolia, 2012). Vision does not normally allow for multiple identities, thus accompanying sound with image remains problematic in my work.

On the other hand, contemporary visual algorithms, such as Google's Deep Dream, comes close to a visual equivalent of auditory ambiguity. Deep Dream is a computer vision program which uses neural networks that are inspired by organization of visual cortex (Mordvintsev, 2016). Over-processed image from Deep Dream creates visual pareidolia as animals and everyday objects are falsly pattern-matched to the parts of the image and drawn on top of input to create a pareidolic effect (Figure 1). If a video is processed instead of a still picture, the constitution of associated images changes on a per-frame basis as objects change their perspective (the algorithm does not involve a spatial model behind the image). Shapes of animals, household tools and parts of landscape flash into existence to match suggestions in the input material. The result is similar to temporary formation of sound object when listening to abstract material and could be explored together with sound.

1.5 Object: unit of analysis

I would like to draw a parallel between the way how sound object develops in Shaeffer's journals and stages of my work. It is possible to see evolution of the concept of sound object as a successive abstraction. The first stage is related to sample based work, which deals with abstraction related to continuity and separation between sound samples. The step was reflected in my desire to develop tools to transition between samples in a fluent way – as a complex, time-variant spectral morphing. Secondly, a concept of discreteness and completeness is further expanded by using synthesis rather than recording. A structural definition of sound object is created, with inspiration from "object" in programming. In object oriented programming paradigm an object is defined by two elements: a data structure, which encodes the state of the object, and an interface of operations, which allow interacting and transforming the data structure. (The Java Tutorials, 2015) Similarly, programming-inspired sound object preserves the state of sound in its data structure and has a number of actions, which can transform it. This concept is further explored in IOIO live interface. Finally, Polyphone deals with proliferation of sound object by allowing to create similar but different versions of it. In parallel to Shaeffer's objet sonore, a variable set of sounds is generated, which, as a whole, would allow identification of essential qualities of sound.

2 Composition

2.1 Spectromorphology as a way to connect perception to synthesis

A discussion of the perceived sound objects and their spatial position requires a vocabulary with spatial and spectral descriptors. Spectromorphological theory, originally coined by Denis Smalley, discusses acousmatic composition from spatial perspective. The basic unit of analysis is spectromorphology, describing *"components of the sound spectrum and how they are shaped through time"* (Smalley, 2014). The first part of the definition includes the representation of the sound in the frequency domain. The second part is concerned with the form of sound as it develops through time.

As Smalley notes, *"The term spectromorphology is not intended as an objective scientific concept – I am not so interested in using literal acoustic analyses of sound, although, of course, these can help us probe the internal details of spectromorphologies. I am much more interested in sound as apprehended by the ear, the changes in spectral energies as heard."* (ibid.) Nevertheless, the theory of spectromorphology discusses sound as spectral energies – representation of spectrum is used. Distinguishing between two types of representations explains the subjective aspect of spectromorphology: representation of sound as it is perceived is different from representation of sound as it is implemented in a program. Using the term spectromorphology as a way to describe perceived sonorities is different from using the description as a synthesis model, at least to some degree.

Even if it may not be possible to capture all the behaviour in a synthesis model based on a spectromorphological description due to practical reasons, using the description to distinguish structural parts is still beneficial. Spectromorphology as a concept bridges the difference between a subjective experience and representational model in a computer program by talking about specific regions in spectrum as if they were specific regions in perceived or imagined space.

2.2 Describing position of sound in space

The experience of space can be discussed as three cross-sections – perspectival space, spectral space and source-bonded space:

- Perspectival space captures the spatial position of the sound on the plane where the listener's vantage point is located. From the perspective of the listener, space in front is prioritized over space behind or to the sides. Changing source position changes the perspective angle of the experienced sound.
- Spectral space describes the spectral qualities of the sound, which contribute to the perceived vertical position.
- Source-bonded space describes the perceived space, as it is affected by the association of sound to specific source. It is expected that sounds of certain nature would occupy specific position and shape of space (e.g. running water may be expected to be located lower in the imagined space and take up an elongated form).

As defined by Smalley, source-bonding is the *"natural tendency to relate sounds to supposed sources and causes or to relate sounds to each other because they have shared or associated origins."* (ibid.) Four factors influence the degree of source-bonding:

- Ability to decode from spectromorphological attributes. Sounds are defined by the change of spectrum through time.
- Everyday experience of sound creates associations to other sounds. Experience of sounds in their normal sonic context leads to the whole unit being identified with better fidelity. Recognition of human agency also contributes to bonding.
- Musical context can also create associations and lead to sounds being identified as having a specific source.

Smalley talks about source-bonding to material world sources, but bonding can happen if an abstract sound shares some qualities of material sound. It is also possible to create local source-bonding within a piece, through association. Thus source-bonding can be seen as externally as well as internally influenced grouping of sound. Judging by the proximity of sounds in time, spectral and perspectival space listener may create an image of specific source.

2.3 Sound object compared to spectromorphology

The property of sounds to be perceived in groups as determined by the proximity in imagined space forms the basis for the concept of sound object as persistent identity. Behavior of sounds is the main factor in creating spatial regions with retained identity: *"Whether the listener can decode the sense of spatiality, from how or where the spectromorphologies are deployed in spectral space, depends on their motion and behavior and how they relate to one another over stretches of time"* (ibid.)

While spectromorphology is concerned with an instance of sound (a specific sound as it is experienced) and its representation as a changing spectrum, the concept of sound object as used in the thesis deals with the set of spectromorphologies and the allowances their structural relationship create in adjoining objects. The structural parts are vaguely defined by their exact spectral occupancy – they

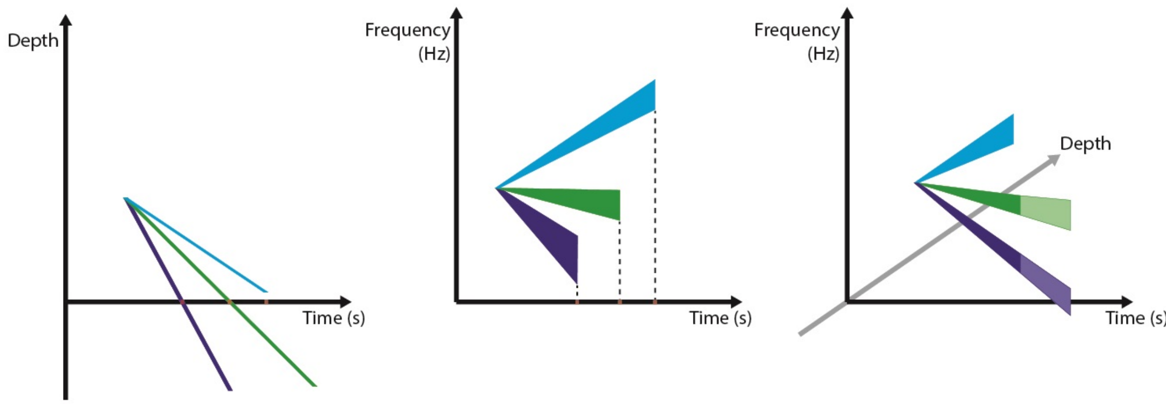


Figure 2: Graphical score of a simple spectromorphology - top, front and perspective views

mainly retain identity across variations by keeping their starting state, behavior and final state. An object is still an instance defined by its data structure, but the transformations it can undergo define a set of potential states it can occupy. Thus thinking about sound object and its allowances involves thinking about a group of related sounds.

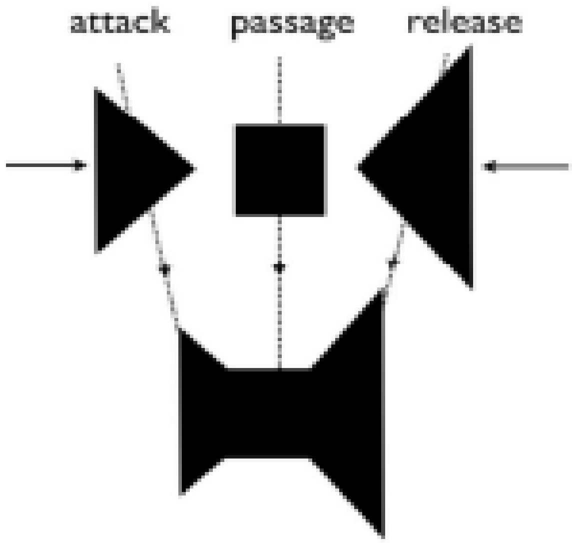
2.4 Sound object: experienced space compared to graphical score

> "The metaphor of an image represents an idea that a number of factors collect and contribute to create and reveal a spatially based form as the music progresses through time" (ibid).

One of the strengths of conceptualizing sound as spatial objects is its relation to a graphical score. Traditonal western music notation can be seen as a quantized graph of frequency over time, with special indications for relative amplitude of notes. Two dimensional graphical score can be used to indicate components of spectromorphology, as shown in Figure 3, left (Blackburn, 2009). A graphical score can be extended to be three dimensional, to include z-axis and visual texture indicating parameters of musical texture (Figure 2, perspective view). Such score captures spectral space and part of perspectival space (proximity to the listener), but does not notate spatialization movement. There is a limited number of dimensions one can visualize and display in static or moving image.

A spatialization diagram with added notation for the type of texture is closer to the experience of sound object. In a spatialization diagram (Figure 3, right) time does not fall onto any of the axes, but manifests as a trajectory of sound object. As a set of perceived components change their position in three dimensions, a spectromorphology is drawn as a volume occupying space. In my experience, it is closest to the mental imagery I attribute to sound object, especially when the different visual texture for each trajectory is used to distinguishing between parts of object.

A drawback to this notation arises in cases when sound movement is complex and the trajectory overlaps with itself, thus creating a superimposed image. Unless a moving image is used, visualizing long periods in such way is not viable and a new "scene" has to be made. The duration a single spatialization graph can represent depends on the amount of change the sound object undergoes.



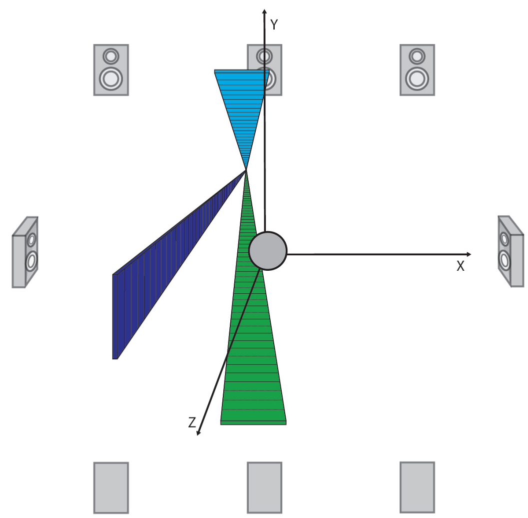


Figure 3: Left: Sound unit construction (Blackburn, 2009) Right: Spatial perspective diagram of a simple spectromorphology

2.5 Time and sound object

How much can be perceived visually in a scene could be related to Stockhauzen's concept of moment: *"When certain characteristics remain constant for a while, a moment is going on. In musical terms, when sounds occur in a certain region, certain register, or in a region of certain dynamics or of certain average speed or tempo, then these characteristics determine the moment".* Similar reasoning is expressed by Smalley: *"A listener needs time to progress from an initial listening encounter with the soundscape to a state of engaging actively and fully in scanning and exploring the spectromorphological and spatial properties on offer. I cannot listen to everything simultaneously and need to devote attention to each of the zones in turn, accumulating a global view, which emerges over time"* (Smalley, 2007). Both authors talk about attentional basis of the duration of a single unit in composition – either influenced by the number of events one can attend to or by amount of time it takes to "scan" each item of interest. From this perspective, sound object is defined by a set of structures or behaviors, which are replaced during variation to such degree, that the listener can still perceive the whole unit in one moment. Thus timing is determined perceptually, by feeling the passage of time through events. Description of Smalley's piece Valley Flow (1992) mentions how the definition of time relates to lack of change and space *"Graduated continuant, or sustained strands and layers, either high or low are very useful in suggesting spatious expansions and over longer periods of time can take on a certain timelessness"* (Smalley, op. cit).

I would like to discuss two works which represent the direction of my compositional practice: Medi (2013, 2:00) and Eigengrau (2016, 11:00). Manifestation of space, separation from object, interaction between space and object and interaction between objects is discussed.

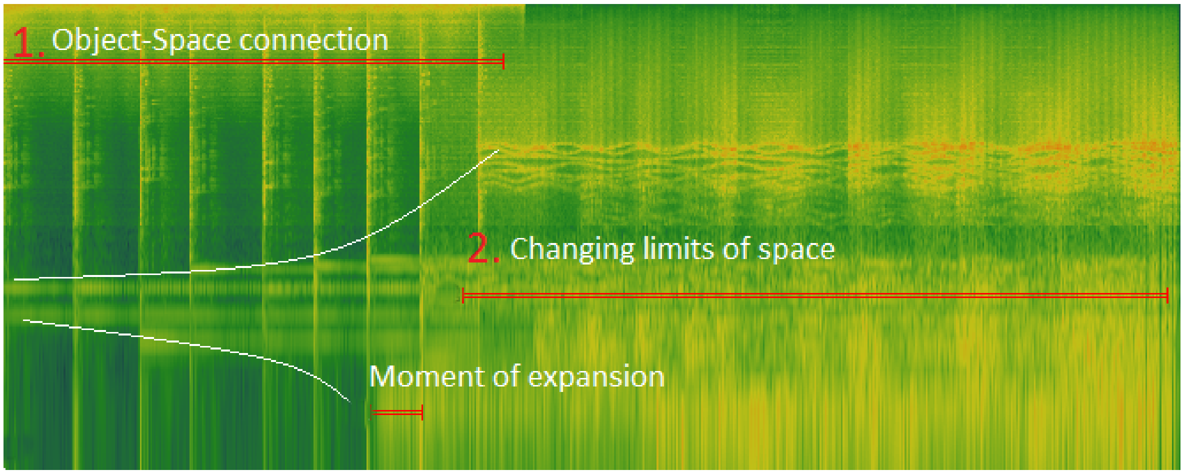


Figure 4: Spectral progression of Medi

2.6 Medi (02:00, 2013)

Medi is a short etude I completed in the first year as I was learning to process sound in SuperCollider. It is not supposed to be a self-contained musical piece, yet I chose to discuss it here since it was the first musical output concerned with space. Reinterpreting it from the spatial perspective clarifies decisions I made intuitively. The etude contains two distinct parts: part one includes discrete impact sounds of breaking glass, part two is diffused multiple layer texture which is modulated to create perception of approaching and receding surface (Figure 4). The effect can best be heard in a very dry listening environment or with headphones.

The role of Medi was to experiment with blending and extension of the background from strongly source-bound sounds. The components which can be heard in the beginning (Figure 4, Object-Space connection) are separated into proximate and distal sounds:

- In the foreground, a surface of banded noise is created, occupying higher regions of perceived space.
- 9 consecutive impact sounds are varying in their material density, positioning relative to the listener and resonance trails.
- Resonances of each impact sound have a role of extending space, as well as connecting the foreground impact sound to the background via intensity modulation. As new resonant trails are added in higher and lower frequencies, spectral space is expanded and the resonances of impact sounds integrate into space.

The second part is marked by a transition to a large space (Figure 4, moment of expansion), created by spectral width and slow modulation rate (influencing the length of perceived moment). Position of the listener changes as slowly modulating components bring the perceived surface forward and backward:

- Resonances with highest (most salient) frequency form the exterior of the surface.
- Stereo field width controls the perspective of the listener.
- Reverberation also controls perceived distance between the listener and the surface.

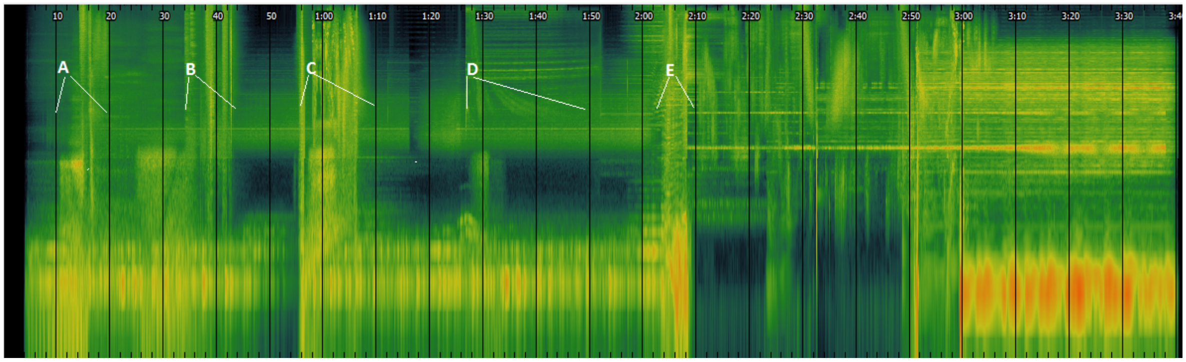


Figure 5: Spectral development in Eigengrau, part I (0:00 - 03:00)

- As three modulations align, the perceived distance is the smallest.

The etude deals with manifestation of space, appearance of entities which are separate from space (created and affecting space) and the effect of changing perspective from the listeners point of view. A simplistic sound object is an actual object (glass bulb), which is varied by changing its core density and pitch of the resonance. The space is affected by the object as the impact sounds interact with the noisy surface. While the piece uses a lot of stereo panning modulation, the effect does not contribute to the angle at which the listener perceives the sound object. Modulation rate is more associated with material properties of the object due to a fast rate pf change. Thus the object is more or less always positioned in front of the listener and only change in its material properties and distance to the listener.

2.7 Eigengrau (11:00, 2016)

"Eigengrau" is a term from early perceptional studies denoting the intrinsic noise people report seeing in absence of visual light. "Eigengrau" or "eigenlicht" (German for intrinsic gray or intrinsic light) alludes to the property of retina to produce self-noise which is indistinguishable from excitation by photons (Barlow, 1972). The concept of intrinsic noise puts the observer in ambiguous situation, where no distinction between what is perceived and what is imagined can be made. The focus of the observer is concentrated inwards, into the image which arises from the machinery of perception itself. From spatial perspective, the piece explores the manifestation of object in the imagined space. Key concepts of the piece are: manifestation of object, interaction of object with space, effect of perspective and morphology of objects.

The introduction of the piece (roughly 0:00 – 02:00) establishes a backdrop of resonant drone with emphasis on lower harmonics, suggesting a large space. The boundaries of the space are reactive to the four events (Figure 5, A-00:10, B-00:35, C-00:55, D-01:25), which have a tendency to emerge and submerge back to the surface – leave the boundaries of perceived space. Each event is developed into a sound object through variation later in the piece.

Depth-related motion is parametrized by amplitude changes, reverberation and granular parameters (i.e. grain smearing in stereo field). Granular parameters also distinguish material properties of the components of objects. One reason to choose granular synthesis is the ability to create material which can "deform" easily. Transition between "states of matter" of an object can happen gradually or abruptly – objects which were perceived to be uniform and static can break into parts with perturbations characteristic of matter cracking, boiling, exploding. All of those state changes deal with an energy transfer and a resulting vector of sound movement. Thus movement of sound is also expressed through change of material properties – motion is associated with transition between solid, liquid and gaseous states. The degree of blending of sound with the backdrop and related material change controls depth movement. Depth movement is also explored in Event E, which can be seen as collapse of space or penetrating the boundary of space to exit to a new space.

Separate events deal with movement on different axes. Horizontal movement is explored in Events A and B, which has a role of establishing the breadth of space. Events C and D explore vertical spatial movement through glissandi of wide noise bands. Since a sound object is defined by repeated behaviour and structure, Events A and B merge together in the second part to create object defined by horizontal movement, while C and D create an object defined by vertical movement.

Musical form is a very broad subject, but definitions of it usually include the relationship between structure and development of material. Generally defined as *"The structure and design of a composition"* (Kennedy and Bourne, 1994) or in a more elaborate way as *"A series of strategies designed to find a successful mean between the opposite extremes of unrelieved repetition and unrelieved alteration"* (Latham, 2002), form in Eigengrau arises as the building block, the sound object, aims to express different aspects of space – space becomes driving parameter of form of the piece. As Smalley comments in his lecture: *"Spatially based forms and space as form <...> are key in understanding the expressiveness of acousmatic image"* (Smalley, op. cit).

2.8 Conclusions about compositional practice

Creating acousmatic music with spatial focus defined concurrent goals in my workflow:

1. Defining perceived image of a sound object which would be holistically interesting from the listeners perspective.

2. Define a structure which is open to further development. One way to achieve that is by defining sound object in separate parts, which retain identity through behavior but are variable.

3. Create a representation of sound object with the tools that are available. A good representation models the perceived sound object, as discussed in next chapter.

4. Develop form which is guided by the properties of the sound object (i.e. change between the focus on breadth and width of space as expressed in the second part of Eigengrau)

Experiencing sounds as they manifest in the same location and at the same time lead to the concept of experienced sound object. Thinking about the physical structure which could relate grouped sounds, the transfer of energy inside the structure, material changes which could occur due to the transfer, lead to the concept of sound object as a compositional unit. As the concept of sound object was developing I was interested in making programs to investigate different aspects of space. In the next chapter I would like to discuss the development of tools as a practical way to clarify concepts about spatiality and sound object.

3 Tool-making

A continuous topic in this thesis is also making of compositional tools. The inspiration for programs often followed a desire to imagine sound visually to expose it to sonic manipulations which could be applied or mapped on a two dimensional surface. Thus the programs are discussed from the standpoint of the spectromorphologies being cut along different axes:

- Y-axis, representing frequency and vertical position in imagined space.
- X-axis, representing development of spectromorphology over time.
- Z-axis, representing depth in the imagined space and dealing mostly with the aspects which separate identities of sound objects occurring at the same time.

The programs are discussed as follows:

- Morpho, an interface to draw development of spectral morphing between two sounds has an emphasis on the frequency parameter, and so it cuts vertically along the Y-axis.
- IOIO, a live interface for triggering manipulations of predefined sounds focuses on intuitive timing – cutting along the X-axis.
- Polyphone, an interface for specifying events as an interpolation between states, focuses on development in multiple parallel components of sound object, thus separating them by behaviour or depth-positioning from the perspective of the listener – thus cutting along Z-axis.

3.1 Sound Object and programs

When Medi (2013) and Eigengrau (2016) were being composed, the term "sound object" referred to different concepts in my mind. In the Listening chapter the pieces are nevertheless discussed using the latest interpretation – sound object as a discrete structural unit, which has a perceptual grounding and a divisible structure, both of which limit and guide connection to other units. When I was making the pieces, I did not conceptualize this way – so the Listening chapter aims to illustrate the power of changed concept in developing ideas which were once considered complete. In Tool-Making I try to show how solving technical problems guided the development of the concept. Thus "sound object" is different for each program discussed in this chapter.

In Morpho sound object is only the mental representation of what is heard, whereas inside the application the sound is divided into structural parts only superficially. In IOIO, sound object is more extensively modeled in the program, yet it only denotes an instance of sound rather than a class (category of sounds). In PolyPhone, the goal is to organize the variety of instances of sounds by exploring the methods used to generate them. Thus PolyPhone is designed to explore classification of sounds.

3.2 Morpho

3.2.1 Context

During my first year in Sonology, I learned about frequency domain representation of sounds transformed with FFT. The conceptual model of sound changed from something which is only divisible in time to an entity which is divisible both in time (horizontally) and in frequency (vertically). Representation of spectrogram became a familiar visual tool to grasp the structure and development of sound in time. At the time my main compositional unit was a sample rather than a parametric representation of synthesis process, therefore I felt that the means to transition from one sound to another were limited to cross-fade. The desire to be able to control the spectral characteristics in time encouraged me to experiment with new tools dealing with frequency representation of sound.

3.2.2 Concept of Morpho

Morpho is a GUI-based tool to create a detailed crossfade from one point in sound file to another point (Figure 6, top row) The app was meant as a convenient way to work with a large sample (e.g. a 10min of synthesis output) to remove unnecessary material and bring relevant events together. The transition between two ends was not to be a crossfade – instead a user could specify the spectromorphological development of one end, while the other end would take the inverse form. The process is better explained in step-wise manner:

1. User listens through the sample for a sound which has an interesting beginning (Sound A) and another sound which has an interesting ending (Sound B).

2. User specifies 2 cut-points at the center of each sound. Duration of the transition is also specified.

3. User visually draws how the continuant develops in time as a spectromorphology. This is done by filling a grid, where x-axis represents duration of continuant and y-axis represents the frequency bin as analyzed with FFT (Figure 6, rows 2 & 3). Each cell in the grid can be either black or white: black cell means that the frequency bin will have amplitude value from the first sound and white cell means that amplitude value will be taken from the second sound. Only black cells are drawn, thus only the development of the first sound is shaped – the second sound fills in the gaps in the spectrum by taking the negative of the grid.

4. The transition can now be rendered to a soundfile.

To clarify the concept, it is possible to describe a simple spectral crossfade in the same visual terminology. A crossfade would mean that all cells in the graph are gradually changing from black to white as the amplitude multiplier for each bin changes. Also the change is equal across all frequency bins (Figure 7). The main difference is that Morpho works by specifying which frequencies to include or exclude at any given time, while crossfade would take all the frequencies and include them from both sounds in a gradually changing proportion.

3.2.3 Functions of Morpho

Implementation of Morpho was completed in Max (Figure 8). Upper part of the application displays the sample, which is looped by default. Yellow field is the selection for transition, with beginning and end point marking cuts. Duration and type of transition and playback control are specified in middle part. The lower part shows three example transitions: leftmost is equivalent to a steep low-pass / high-pass filter sweep, middle takes even and odd bins from each sound and has no development in time, right shows a spectromorphology with material from sound B in higher frequencies gradually being removed.

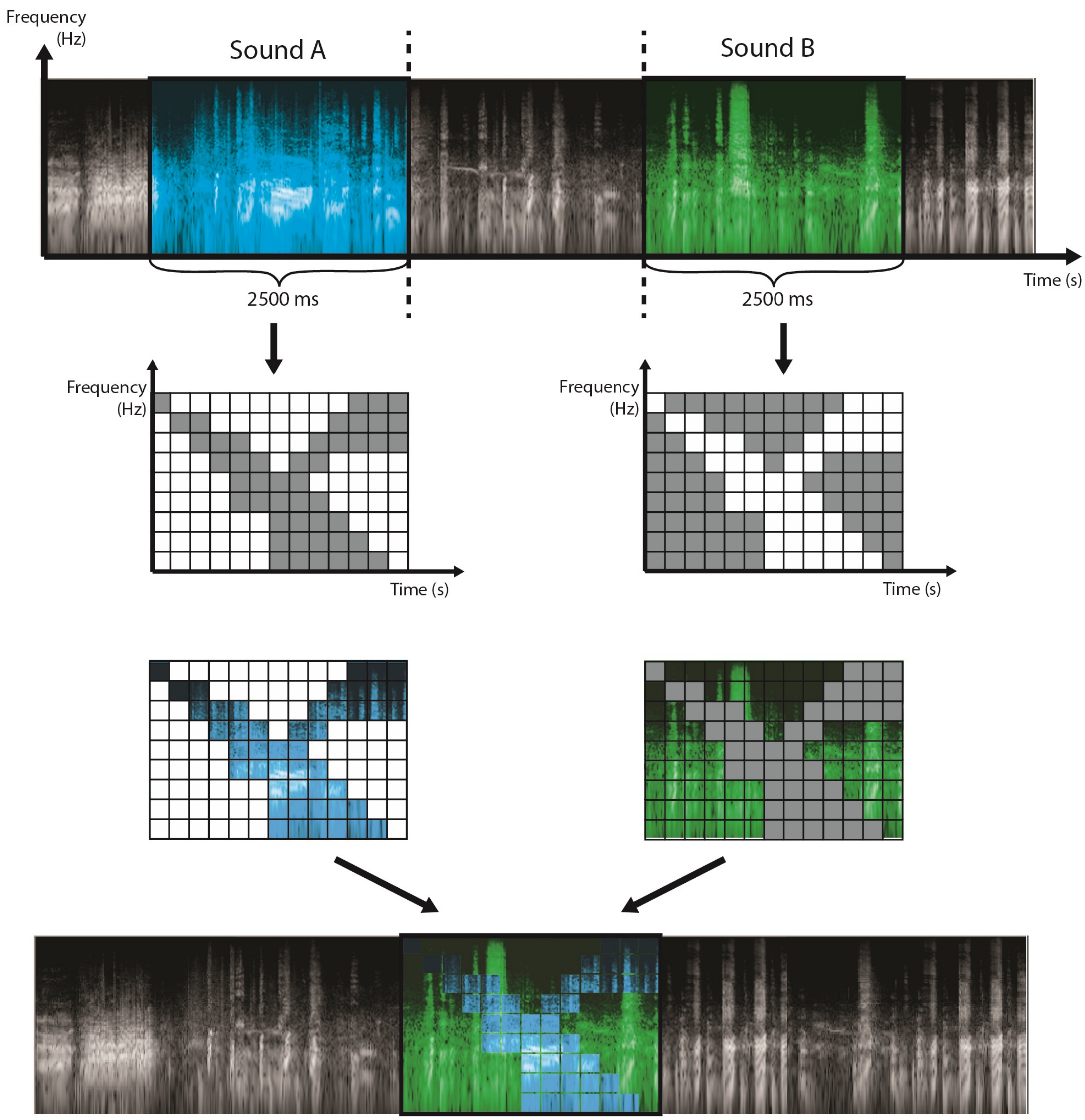


Figure 6: Model behind Morpho - creating spectromorphological transition

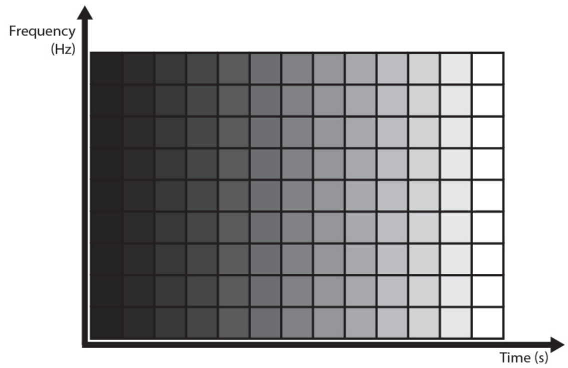


Figure 7: Representation of a simple crossfade in visual terms

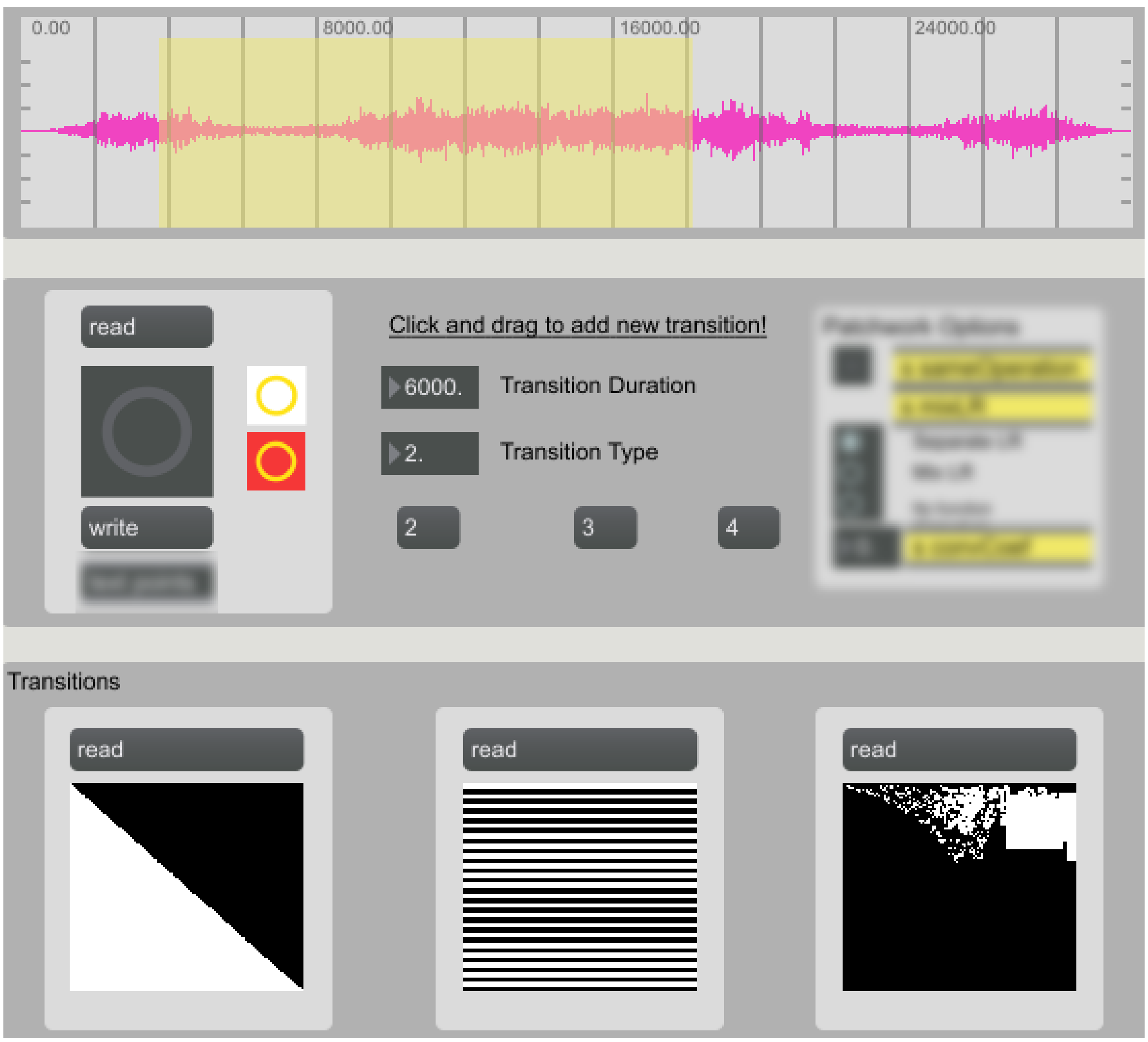


Figure 8: GUI of Morpho (debugging interface blurred-out for clarity)

3.2.4 Evaluating results of Morpho

A program can be seen as an effort to investigate different development during spectromorphologies as a way to investigate how the recorded sounds are internally shaped – to discover generalities among material sounds. The big assumption in this case was that the right shape of the "cutout" mask should have divided the sound into different structural parts of perceived sound object. Once parts were separated, it would be possible to find two sounds which combine as well as the revealed parts allow.

Thinking about sound in the frequency domain was a good exercise and a step into the direction of controlling the development of timbre over time. The strength of combining parts of spectra rely on the representation of both sounds as Fourier transformed data. The issue turned out to be mainly the mismatch between what is perceived to be the sound-object in its audio-visual form and how the representation in the program divides the sound. I imagined that creating a detailed and time-variant filter (as the transition grid can be seen to do) would allow to isolate elements with separate identities. A naive example would be recording a bang on a dried branch, where low thud would be the resonance in the thickest part of the branch, and high hiss would be the smaller branches and dried leaves moving. Sculpting out the spectromorphology would allow to hear the parts of the tree as if they were recorded separately. This of course was not true and although impacts to a heavy branch and a light branch have more and less lower frequencies, the identity of the sound as we perceive it depends on full spectrum. What I wanted to divide was physical structure of the sound source or resulting different behavior in different parts of the sound, but what got divided with the application was the representation of frequencies rather than behaviors.

Perhaps a better approach would have been separating the transient parts of the sound from continuants, since analysis allows to separate noisy parts from resonances. That would be an example, which successfully divides the sound object structurally (resonant body from the location sounding on immediate impact), as it can separate different behaviors.

The conclusion was that I need to construct a model of sound myself through synthesis or editing for all the parameters to be separate and retain their behavior. The conceptual model at the basis of the program did not fit the intention of the program and, even though I found uses for Morpho with synthetic material, the need for constructing the sound object from its components became apparent.

3.3 IOIO

3.3.1 Context

IOIO was conceived as an interface for live control of transformations on the sound object and implemented in SuperCollider. The concept was to bridge the listening and composition process with the following reasoning:

1. experiencing (or imagining) sound involves creating a mental representation of it as an object;
2. programmed sound object should be based on the mental representation, with manipulations having equivalents between the imagined and modeled sound object.
3. It should be possible to manipulate the object live, with an aim to match imagined transformation and produced transformation.

The name of the program IOIO – Input Output Input Output – represents two stage role of listening in performance / composition. First Input-Output feedback happens when the object is being modeled.

The goal is to find a transformation on the parameters which represents a material change in object state. This can only be done by verifying it by perception, thus feedback is necessary. Second Input-Output refers to the performance stage. Once a satisfying model of object is created, object is being manipulated live. The change between the states of object – the order of changes, their intensity, their degree is the focus of the performance. The program is the main tool in creating a piece Dreamstate #1 (2015).

3.3.2 Model of sound object

Since the goal of a program was to create a generalized interface for manipulations, all sound objects created for IOIO had to share the same sets of parameters and transformations. Each Sound Object consisted of following parts:

- A SynthDef controlling the synthesis process.
- An interface to map all synthdef parameters to generalized parameters (many-to-8).
- An interface defining the gestures that can be applied to the sound object. Each gesture was expressed through envelopes of the generalized parameters.

The term gesture implies some human agency and intention and can be related to the concept that the manipulation occurs on an object which is imagined to be material like. The manipulation of the object is also observed from the listener's vantage point, often passing into the inside space of the object (by extending the object widely in the listening field). Thus there is an implied invitation to perceive the manipulations as corporeal – on the program side the manipulation is represented with a simple set of envelopes, but the design and concept is to create a manipulation which expresses recognizable agency.

In context of IOIO, the objects differ from each other in the mapping of generalized parameters, synthesis process and envelope set. They are similar to each other in that each is accessible using the same set of parameters and each could be transformed with the set of equally named gestures. In this respect, IOIO is influenced by EPOC (Gunnarson, 2012). Highest order parameters can be divided in the following categories:

- Standard synthesizer parameters:
  - Frequency – to control a tonal center of the texture, implemented either as the fundamental frequency for additive synthesis processes, filter band center or pitch of the grain for granular processes.
  - Amplitude – since objects are continuously sounding, so that the performer always knows their state, amplitude is the only parameter to control when the object starts and finishes.
- Granular synthesis related parameters:
  - Mass – corresponding to the perceived mass of the texture and implemented by duration of the grain. Longer grains filling time line with more densely packed grains.
  - Speed – corresponding to the perceived rate of change in the texture (i.e. how quickly grains follow each other, especially if they are perceived as different and separate).

  - Entropy – corresponding to orderliness of grains and expressed through the degree of periodicity of grains (under the threshold of pitch).

- Global textural modifiers controlling the timbre of sound:
  - Color – Col relates to amount of harmonicity of the sound and harmonic relationships.
  - Surface – related to distortion and roughness of sound.

- Spatial occupancy in multichannel system:
  - Space – corresponding to the dispersion of the sound in multichannel system and related impression of the size of space the object resides in (widely dispersed, decorrelated noise sources creating a large space).
  - Location – this is an extra parameter, which was not implemented inside IOIO, but done at the editing stage. Controls the location of the object in imagined space.

Standard, granular and timbre parameters can be seen as related to the material properties of sound object. Spatial parameters are related to the representation of space the object occupies.

3.3.3 Behavior of sound object

The behavior of sound object is specified through the transformations of generalized parameters. Instead of applying additional synthesis processing to create behavior, the goal with IOIO was to define the behavior in terms of the reduced set of parameters. Behavior was to be defined for each individual object, due to differences in mapping and synthesis process, but the aim was to keep to a fixed set of opposite gestures:

- Drop / Raise – a transformation which is concerned with the position of spectromorphology on the vertical axis, also corresponding to its perceived vertical position in space. Parameters other than frequency control the key transition moments of the object – its start of movement, its arrival and supposed impact, changes in the material and granularity as a result of imagined energy transfer.

- Expand / Contract – a transformation dealing with occupancy of the object in the "visual" field of the listener, as observed from the listeners vantage point.

- Emerge / Submerge – a transformation controlling the relationship between the background and foreground or the depth of the object in the prospective space.

The result was a set of controllable sounds, each having the same interface of structural parameters and behavioral gestures. It was done as an abstraction, to let user focus on more global changes and their role in the final output. The idea was to simplify control to make the experience of using IOIO as close to as possible to manipulation of the imagined object.

3.3.4 The role of intuition

As an instrument or as a compositional tool (if used to generate material) IOIO relies on intuition in two ways:

- Intuitive timing and intensity of the gestures gives the program instrumental quality.
- Matching the sound output to the imagined behavior of audio-visual sound object. In this case the graphical score would be envisioned first and then sounds generated according to it. This is more in line with compositional tool.

It can also be argued, that composition itself, as a process of making decisions about the structures in the piece and their general integration into the form is not within the scope of the program itself. Instead the composition is done when drawing the graphical score as well as designing an object model. Designing an object involves compositional decisions by confining the object to specific material qualities and behaviors through choice of synthesis process and the mapping of internal parameters. Structure of the piece is then governed by the allowances of the object – thus the effect on composition. Yet, the creation of object itself is not facilitated by the program, thus IOIO does not aid composition directly.

3.3.5 Changing concept of sound object

Sound object here denotes an instance of sound – the program may contain one representation of object, defined by its material properties through parameters and its behavior through gestures. Gestures act on the object, changing its state relative to the previous state. The choice to map relative instead of absolute was a deliberate decision. When mapping the parameters, it would be possible to present them in absolute change or relative change. In the first case, parameter is always changed on the absolute scale, in relation to the smallest possible value. For example, a slider or a pot, which have predefined limits suit this setup. The mapping encourages relating specific parts of the slider to specific states of the sound object, thus creating an understanding of state space. The act of adjusting the controller would be equal to moving to a specific part of the space. However in the second case, parameter is changed relative to the previous value. For example, a rotary encoder is better suitable for the setup. The change in the controller sets value relative to the previous state, thus the current state of the object moves into the focus of attention.

The idea of IOIO was to change the object state in steps relative to the previous state. For example, a high velocity "Drop" gesture would create a full-range movement of the object and high energy impact and could be only done once, before a lower parameter limit is reached and an opposite gesture needs to be used. "Raise" could then be applied in small steps, taking several to reach the highest parameter limit. How many steps were available before the limits were reached was an estimation by ear each time. A distinction between an instrument and a control interface can be made on this property – control interface presents a state space and lets you enter data to reach a specific part of it. An instrument requires auditory feedback.

The relative setup did encourage focusing on the sound as well as on the change. It is possible to look at the learning goal of the program to see what domain of music is being explored: in case of absolute mapping it would be the parameter space – the set of possible states the object can occupy; in case of relative mapping it would be the sequence of changes – how a series of same gestures compare when applied from the object at specific position in space. Therefore, relative mapping was a decision to investigate behavior.

3.3.6 Evaluating results of IOIO

However in practice, the setup also felt confusing. One cause for that would be the instrument-like nature that required more practice to predict how the changes would affect the output. The same gesture (corresponding to the same physical action on the interface) would produce different sound dependent on the starting position of the object. The problem is especially pronounced when approaching the min/max of the full range of the parameter, a condition which should be clearly audible to the player, since the state of the object is determined by listening alone. Reaching the "edge" state can happen in two ways. In first case the ranges are set to limit the sound to audible and aesthetically pleasing states, thus the end point is not very pronounced as there is not much difference between one of the states in the middle of parameter ranges and the maxima. That requires much more practice before it could be recognized without error. In the second case the ranges extend to the states which were not intended to be normally heard in the performance – i.e. too loud or too high in pitch. That would indicate the approach of the parameter ranges clearly, yet the listener is then put into the position where a significant part of the performance is navigating away from the states which should not be heard. This may be a problem depending on the amount of practice by the player as well as the intended goal behind a composition.

The purpose of IOIO was to create a situation, where an imagined sound object could be manipulated intuitively and quickly with the aim to explore the behaviors of sound objects in general. After experimenting with the visual presentation of parameters, to make it more clear when the maxima were reached, I decided to remap the parameters in an absolute way. As a result, the focus was shifted towards parameter space and the identity on the object as defined by the location inside the parameter space. The direction is further explored in PolyPhone.

3.4 PolyPhone

3.4.1 Organizing sonic exploration in states

While IOIO approached the live aspect of sound object through predefined gestures, PolyPhone is focused on generating behavior as a transition between states of sound object. As a result, the interface of Polyphone displays representation of state instead of transformations the state can undergo, like in IOIO. A state is defined as a moment that can retain the identity of the sound continuously or can be used as a reference point to define change. Smalley comments on the continuity of the state in the lecture on spectromorphologies: *"It is not that the state is necessarily a stable slice in time, but it is convenient to think of states"* (Smalley 2014).

The purpose of the state is to:

- Create a working unit, which is similar to object-oriented-like object. Once we are able to encode data, only then we can define what transformation the data can undergo. Therefore I was inclined to take a structural standpoint towards sound from the beginning.
- Present the variety of units in some organized manner, which would facilitate the exploration of topology of the sound stemming from that synthesis process.

A state is defined by the following components:

- A synthesis process, defined in SuperCollider SynthDef, accessible via a set of parameters
- A list of values of the parameters.
- Ranges that the potential value parameters can occupy, if changed by randomization, denoting related state space.
- As the name of the program suggests, a set of SynthDef instances as voices, each specifying an independent structural part or behavior in the sound object.

As a result, a state can define a particular sound, usually a texture, which has internal movement, but could most likely be perceived as not changing. State is also related to other states in two ways: top-down, through the nature of linearity of GUI elements and bottom-up, through controlled randomization abilities. The top-down relation arises simply because the parameters are represented with sliders, so the GUI choice in itself encourages linear changes between values that are close. The goal is to use the interface to specify sound I had imagined beforehand. The bottom-up relation represents an opposite way in relating the states – it is possible to randomize the parameters within ranges, but without necessarily paying attention to the specific values of parameters in the parameter space.

Both of the relationships require auditory feedback – the user judges if the state is interesting and will be used in further processes by ear. However, the nature of how a state can be reached determines which states will be discovered further. In other words, since the parameters can encode a very large number of different sounds – a large state space – even small decisions about how the user interacts with the application determine what sounds will be discovered sooner. The process of designing an application such as PolyPhone can be seen as organization of the state space, revealing the topology of the sound object.

At the level of GUI, organization of state space can be seen in the parameter slider screen (Figure 9, top row). In the default "Per Voice" organization, the state is represented by a stack of parameter sliders. If the state has more than one voice, additional voices are presented in the tab strip and can be reached by swiping the screen horizontally. Each tab corresponds to an independent structural part of the sound object. An alternative organization is "Per Parameter", so the sliders are reorganized to display the same parameter from all voices (Figure 11, bottom row).

User can switch between the two representations while working on the sound. The switch represents a different "cut" across the sound object. "Per Voice" perspective separates the sound into structural components, where each corresponds to different behavior or occupancy in the spectrum. For example, a voice can be named "High drone" or "Low crackling". This representation allows contrasting and separating voices to be the focus of GUI. "Per parameter" perspective puts the same parameters of different voices in the same page to compare them. Thus the representation focuses on producing similarities between the voices and is useful to specify a state which represents a global change. It also allows contrasting parameter change between each other, if is the desired effect (i.e. fast change in amplitude, slow change in pitch).

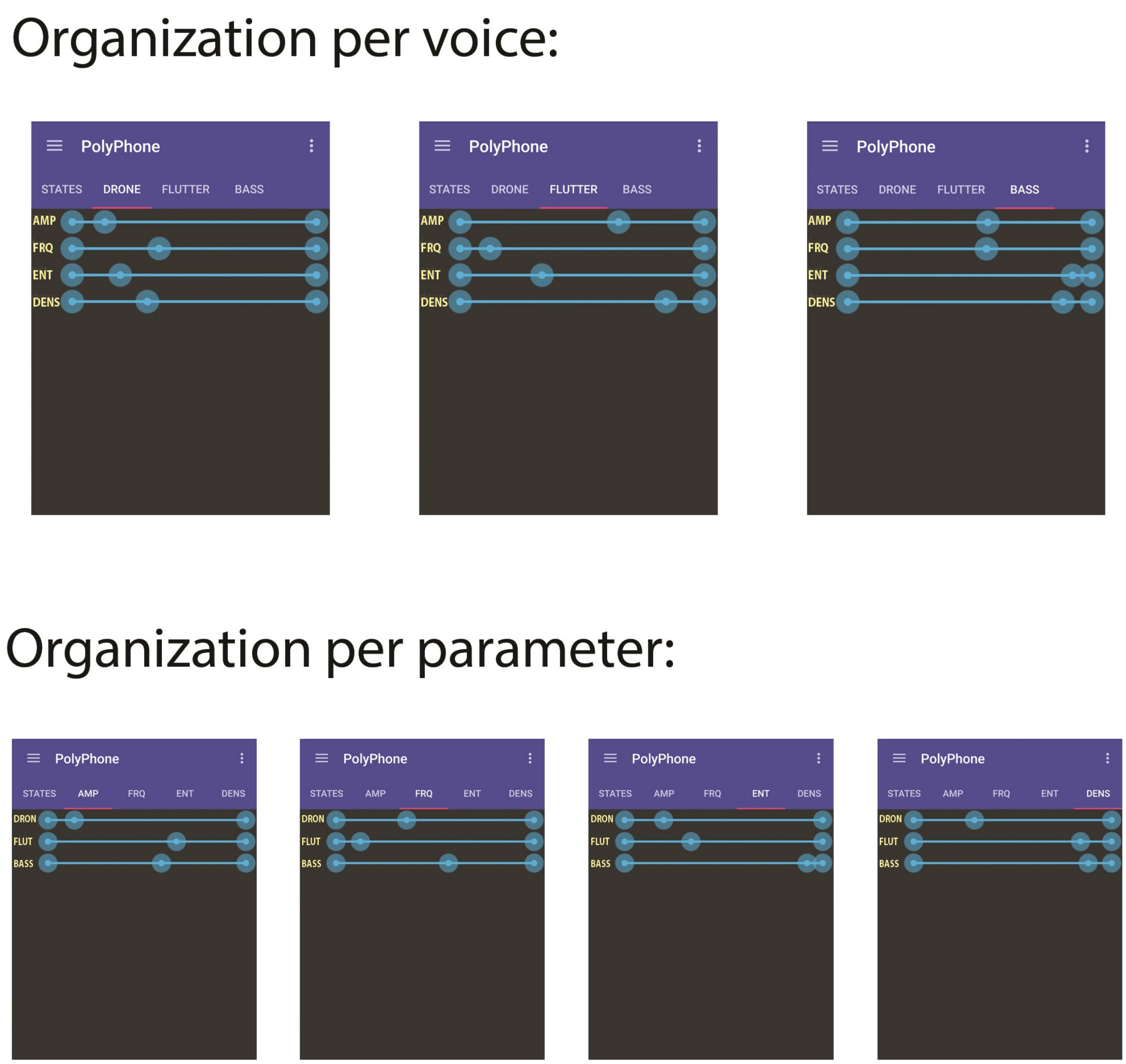


Figure 9: Two representations of the same state in PolyPhone

3.4.2 Traversing State space

With the concepts of states I had the way to explore the possible shapes sound can take, the interest now was in exploring effects that different sequences of states would create. Event was defined simply as a sequence of states, which would be interpolated between each other, creating a multi-parameter, multi-voice envelope set. Creation of an event is represented in Figure 10. The purpose of event was to:

- Create movement in sound, which is still defined by key points – states.
- Control the trajectory between the states, including the ability to choose the the curve between linear, and various degrees of exponential.
- Explore the morphology of movement through the state space.

Event is based on the idea of abstract gesture, creating a shape in space. If the majority of the parameters in the synthesizer are linear and relate to space, then the movement in state space may represent movement in perceived space. The resolution of the trajectory depends on the number of states used in defining the event – it may contain as many states as convenient to display on the phone screen.

The morphology of movement is explored by creating siblings of the event. It is possible to make related trajectories by keeping the beginning and ending states the same, but randomizing the intermediate states. The randomization ranges of the state allow creating events which perceptually belong to the same category – the internal motion is varied while the start/finish is retained.

3.4.3 Controlled randomness

Randomization of state to generate sibling states is one of the powerful aspects of PolyPhone. At the time of programming, I felt that adding the randomization would be additional feature, yet it turned out to be the main driving process to generate new material. The direction in which the state is randomized is determined by 3 factors:

- Random ranges described previously determine the min/max values parameters can take. Conceptually, they should determine the boundaries of parameters for the event.
- Random mask locks particular parameters or voices from randomization (Figure 11) The mask represents a cut through the sound – either by locking a certain parameter (i.e. as you transition between State A and State B, keep the entropy the same) or by structure (i.e. keep the "Flutter" constant, but change the "Bass" component).
- Randomization step size – a global parameter describes the degree to which the randomization is applied. This determines the degree of statistical difference in the state space, but not the direction.

Although PolyPhone does not include any visual representation of state space itself beyond a single state, the modes of traversing the space allow creating families of related states and events. For example, if a small step size (e.g. 1%) is chosen, it is possible to move by small distances. By changing which parts of the state are affected the direction of movement inside the state space is controlled.

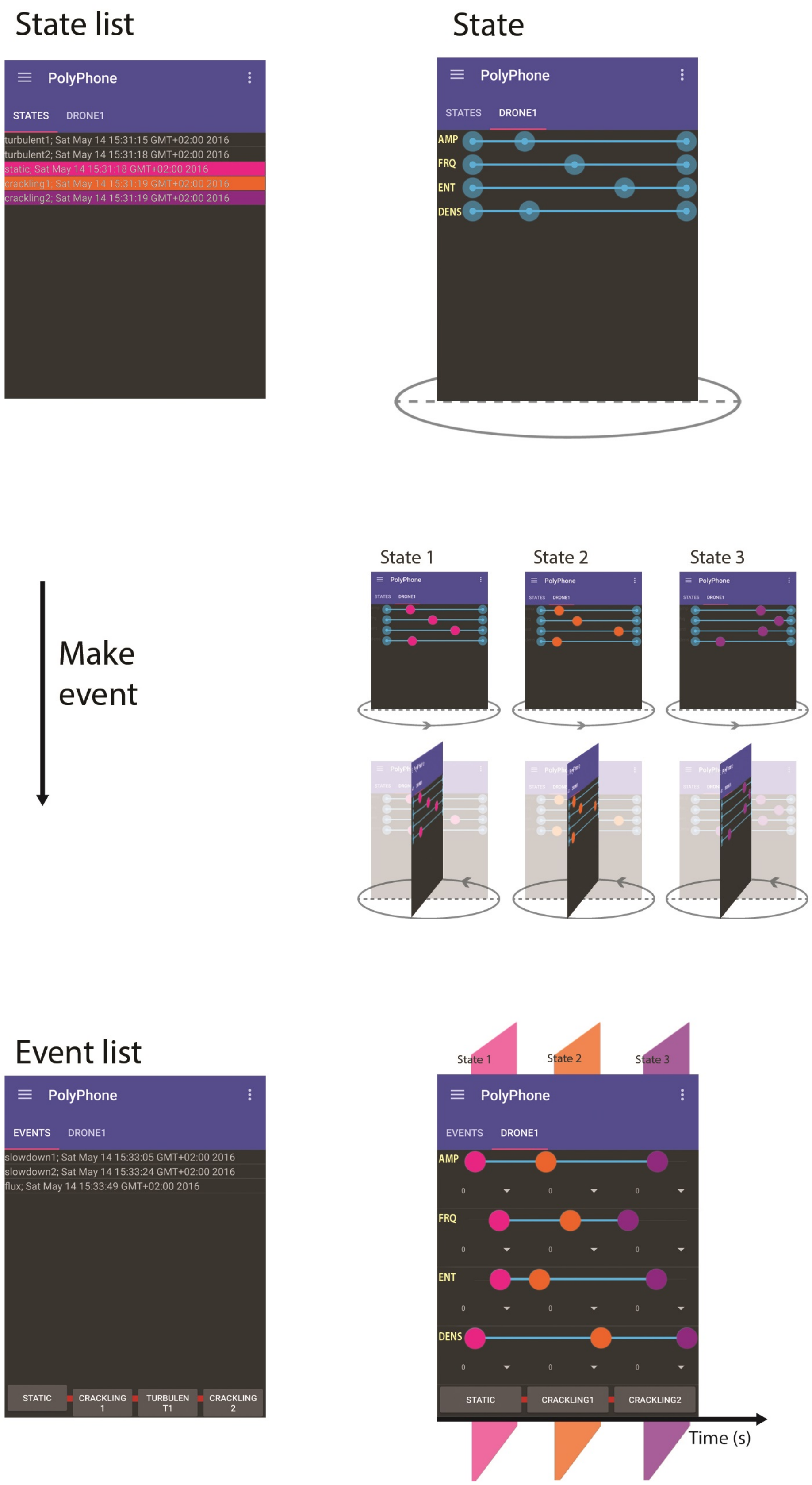


Figure 10: PolyPhone - representation of a state, sequencing of states to make an event

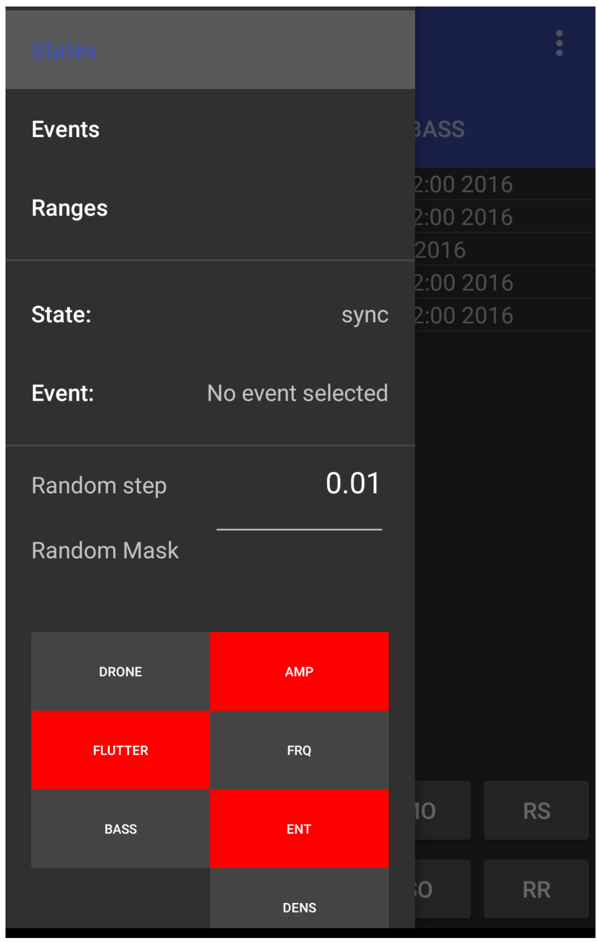


Figure 11: Controlled randomization - red mask shows which parts of state will not be randomized

3.4.4 Behavior of Sound Object: visual vs auditory cues to movement

The idea that spatial movement can be achieved by interpolating between states is based on visual metaphor. Spatial movement of sound object is achieved by parametrically encoding its state at certain positions in space and interpolating between the parameters to form the trajectory of movement. In a graph, the trajectory can be defined by start and finishing points, with any number of points in between to denote its shape.

Visual metaphor holds for sonic parameters which change linearly as the object changes its position in space, such as panning of sound source in multi-speaker array or other parameters which linearly denote the distance of sound to the listener (e.g. low pass filtering, amplitude, reverberation). However, while making pieces I have found that movement is best perceived in a cause-and-effect manner, with a specific event starting the movement. This makes sense from natural perspective, where the movement results from a transfer of energy – something has to cause it.

Not to say that all causes of object movements in the imaginary space have to be external, as that would reduce the choice of sounds to impacts only. Yet, perceiving the cause of the start of movement adds to the impression that the abstract sound has physical restraints. The realization meant that the concept of sound object has to be extended outside the boundaries of Polyphone. The strength of PolyPhone was in creating the trajectories, but not the beginnings / endings of the object movement. As a result it could not capture whole object – its other components needed to be added in sound editor.

3.4.5 Is the structure of PolyPhone limiting the compositional output?

PolyPhone involves presenting the parameters of sound in a linear way, which could be seen as limiting compositional output. The aim of linear organization is in fact to organize and limit the mobility of sound object, so that an intentional exploration can take place. On the other hand, the randomization part of the program ignores the linearity aspect. Thus the answer to this question depends on what we consider to be the focus of creative process and composition. One can see composition as a process of exploration, where the compositional method has a goal of arriving at new exciting sonorities. PolyPhone organizes this process with GUI as well as guided randomization.

3.4.6 State: extending the concept

As I got more familiar with the concept of state, a need to operate on a more universal unit has arisen. State thus can be seen as a container for either a texture or a gesture, more dealing with the organization of related sounds rather than the initial nature of the sound itself. State can be seen as a building block of state space, which could organize gestures, static textures, sections of a piece, transformations etc. Implementing this universality is one of my future interests.

One of the directions would be to extend PolyPhone to include a visual represenation of state space which would present genealogical relationships between states. In PolyPhone new states can be generated by adding a small displacement with positive or negative direction to all parameters. The overall direction of the change is unknown – new state is only statistically different from the previous one. This movement can be called mutation. Figure 12 shows visual representation of parameter space with two different movement types generating new states. The central axis (marked green) represents mutation, which changes state in undefined direction with a small degree. The remaining cells are generated by recombination (combination of two adjacent cells) – combining parameters and voices from adjoining cells. Choosing a state on the mutation axis lets the user pick two states which are different to a specific degree, but without need to know which parameters encode it. Recombination involves combining newly mutated states between each other to create a controlled continuum of sounds. The sonic result of recombinations depends on the rule used to combine two states: is it an interpolation (taking equivalent parameters and averaging them) or discrete mixing (i.e. if two states have 4 voices each, one cell could use two upper voices and one cell two lower voices)? Movement trajectory across the state space creates a development in sound, with predictable start and finish points.

3.4.7 Why Android?

A lot of work in PolyPhone involved learning the basics of app programming for Android. At first sight, PolyPhone does not utilize any of mobile specific advantages in comparison to PC – no sensors such as accelerometer, light sensor, compass or barometer are used. A lot of overhead arises from Android being mobile OS – restrictions in memory, freezing of background processes and exotic separation between GUI and functional part of the program add to the complexity. Nevertheless, there were two reasons to pick a mobile platform: interfacing with a touch screen and mobility. Touchscreen sliders are more versatile to physical sliders since they allow both continuous linear change of parameter by sliding as well as jumping to a specific value if a touch is registered. Since PolyPhone descends from the ideas of live control, some of the parts of the program act as live instrument. For example, during an event, as the interpolation between states is played out, it is possible to grab some of the sliders and override the interpolation. The sliders also support multitouch, so several parameters can in theory be adjusted. The second advantage comes from the fact that PolyPhone is mobile and separate from the synthesis process that runs on a more powerful and stationary server. The focus of my work involves expression of sound objects in space. Therefore perspective of the listener is important in creating realistic movement. PolyPhone lets me move around the space as I adjust the sound object properties to perceive as people sitting in different parts of the concert hall would.

4 Conclusions

After creating and using the applications which were inspired by the conceptual model of sound object, I was facing a question of their identity: does the program express an aspect of conceptual model and, to maintain consistency in the compositional practice, should be used only the way it was intended, or should I deal with the app as a way in which I can transform sound abstractly, without a conceptual model? My intuitive answer was to use it as is, disregarding the original purpose. However, there are reasons not to abandon the original intention completely.

An application is just a mindless tool, allowing a set of predefined symbolic transformations. What those transformations mean to the composer, depends entirely on the type of material and its significance. A set of envelopes in IOIO representing a diffusion of specific, localized sound mass into a cloud-like texture only signify "Diffuse" if both the material and parametric mappings are correct – if they make sense, checked perceptually, as something they were intended to sound. Therefore, the geometry of envelopes for the "Diffuse" gesture does not carry the information of perceptual diffusion – it is the right combination of material, parameter mapping and envelopes.

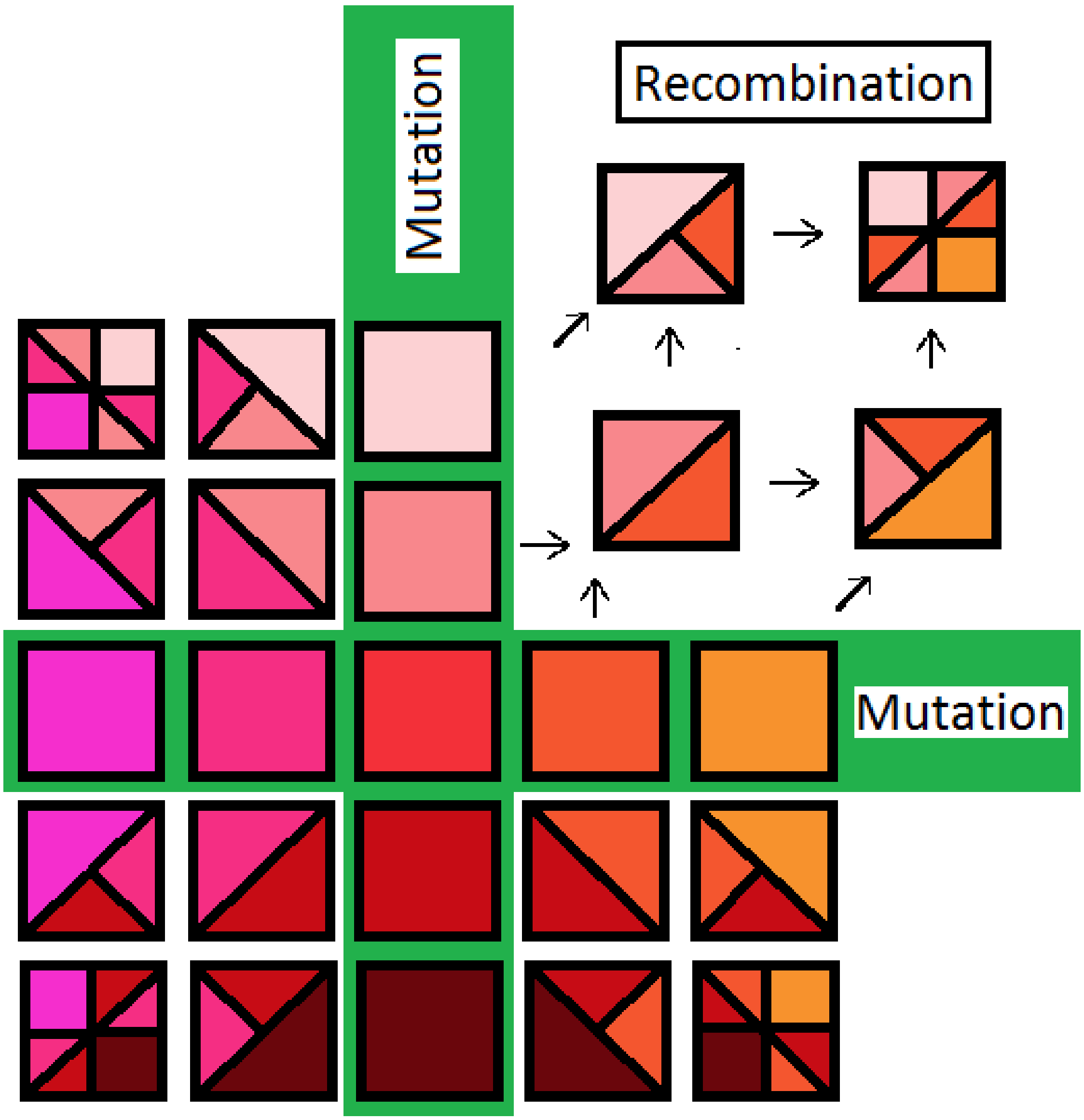


Figure 12: Visual representation of state space

On the other hand, once we have made a single "successful" diffusion gesture, it is reasonable to expect that variations will also carry related perceptual characteristics: similar sound objects, transformed with similar set of envelopes will sound familiar. It is therefore practical to aim for consistent signification when naming transformations. To maintain the identity of the transformation I tried to incorporate this gradual property in all my programs. In Polyphone, it is possible to step-randomize the state to gradually move away from the original. The movement is necessarily randomized, so that the direction of it is not intentional – it is only statistically guaranteed that sound state B will sound different than sound state A, when its parameters have undergone a sufficient number of transformations. In IOIO the graduality is controlled by the velocity with which the transformation button is pressed. In Morpho, it is possible to draw similar masks to keep the transitions similar. Yet, all three programs allow for transformations which change the sound dramatically, changing the original identity and label of transformation, thus transgressing the model.

My conclusion is that maintaining a conceptual model as the composition of a piece progresses acts as a unifying factor, whereas treating the transformation as is, without any presupposition acts as diversifying factor. Both situations allow proliferation of material in different ways: in maintaining conceptual model new relationships between components become apparent, in discarding the model new ways of using the tools are promoted, which as they accumulate, can lead to new conceptual models. For example, thinking about the concept of sound object from the perspective of spatiality gave rise to new directions: Structural – e.g. what morphological differences can be made while keeping the identity of sound object? what would it sound to extend the object with vertical strands at the continuant stage? Behavioral – e.g. what does a rotation of the object sound like? What structural elements need to be changed to create impression of movement across vertical space.

Similarly, once the concept of space was added, more concepts arose in form of questions: Structural – e.g. can different shapes of space be created? Can space exist inside another space? Behavioral – e.g. what is the interaction of space and object? How do two objects interact (since this deals with two volumetric identities, they are by definition situated in some kind of an extended medium – space)?

These questions opened up new directions in my works and improved on the intentionality aspect. Similarly, PolyPhone was designed with these concepts in mind. However, once I started using the app more, some of the functions turned out to be less useful and others gained more significance.

When making PolyPhone, I was thinking about a sound object as a set of states, where each state "pins down" the object in space (in terms of structural components and instantaneous behavior), whereas the "areas" between the states (period between each state) were left for the program to fill in automatically. The application was oriented in making it easy to specify the general structural contour of the sound with a simple thought process: "We need sound A, sound B and sound C in the following order, spaced at these distances in time". However, once the application was made, several problems with the concept arose.

Firstly, it became apparent that the behavior the interpolation creates between states is often more characteristic of the identity of sound that the sequence of states. Therefore, the main idea of specifying each part of the sound in top down manner was no longer fully feasible – I found myself using the parameter slider screen less and simply ignoring what exact value I entered for each parameter. The problem arose partially because of mismatch between non-linear parameters in SynthDef (e.g. reading across a heterogenous buffer for a granular process) and linear presentation in the interface – a slider. As D'Scipio comments on modelling sound based on granular synthesis: *"The design task* *cannot be accomplished in a goal-driven style of design – at least not before the composer has observed the behavior of the micro-level process extensively enough to predict its outcomes. The overall shape of the final sound object is more operationalized than produced following a sound image which pre-exists to the composition"* (Di Scipio, 1994). The rationale for using interpolation between states was to fill up the predictable part of sound object automatically, which largely relied on the linearity of perceived sounds between two states. Large jumps in non-linear parameters created behavior which was too erratic, changing the identity of the sound.

Secondly, linearity of sliders turned out to be a limitation in other ways (even though it was necessary for the initial concept to work). Intuitively specifying the values with sliders turned out to be a creatively limiting experience – I felt the need to get a new sound suddenly, which would somehow be similar to the experience of the listener hearing that sound as a sudden onset. Instead, working the sound parameter by parameter to incrementally reach the desired state lead to conditions where the sound is so familiar it is no longer surprising. Also, intuition posed a problem: some new sounds could not be imagined by intuition alone, while some imagined sounds did not map to existing parameters intuitively.

Both of the problems drew my attention away from what I imagined would be the main interface of Polyphone – the parameter sliders – to the parts which controlled randomization. A more useful approach emerged: firstly, a base state for the sound object was found by randomizing all the parameters in full ranges until something sounded interesting. Secondly, a base state was used to generate child states, which differed from each other only slightly. This was done with step randomization, changing only some parameters by small amount.

Since several voices were involved and each had multiple parameters, usually the sound would have a component which needed adjusting – so the parameter sliders were still used. However, the concept of the exploration of parameter space in Polyphone was now different: instead of connecting new unrelated states with interpolation, the app was used to find new "islands" of related sounds, where movement withing the region would create behaviour defining the state (include image reference). Therefore the concept of sound object became different – it changed its emphasis from identity through structure of a state to behavior of transitioning between states. A sound object would be different from other objects if the difference between the regions in parameter space was large enough, so that the behavior changes to a sufficient degree. Permutation of state order in the event creation process created different behaviour, resulting in similarly sounding but different objects.

To review the partial quotation of Koenig at the beginning of the thesis, *"Aesthetics deal with perception, and thus more with the listener than the composer, who, however, during discussions with colleagues and listeners, when reading criticisms of concerts, listening to his own works, is within a feedback circuit, so that perception, both his own and that of others, affects his composing. This causes the aesthetic experience to be transformed into the rules of compositional craft. <...> Discussion of aesthetic questions can only provide stimulation, for there is no cut-and-dried system of instruction for electronic music."* (Koenig, 1968). An answer to an aesthetic question will always include perception in the steps one takes. There seems to be no other direct way between manipulating representation of sound and creating quality music – one has to listen, perceive and discover what works. Composer is always a listener. Yet leaning to either side creates two different approaches: first one requires analysis of the subjective experience and its grounding in the perception science to make a model of sound; second puts emphasis on composition as an independent process, dealing with the ways structures can be generated and related. Accordingly, all the tools I have created aim to reduce the gap between applying a symbolic transformation and hearing its results, thus taking the first approach. At the same time, the balance between listening role and composing role seems to be shifting to the right.

In the beginning my approach in composing acousmatic music was to find the best model to represent listening experience. The right representation would be based on assumptions about perception (multimodal integration into spectromorphologies) and perceived structure (sounds as objects). It would also allow manipulation of sound in real time. The ideal goal was to generate experience as directly as possible – by applying transformations on the sound object in the program as one would imagine the perceived sound object could be transformed. This approach was useful in two ways: it let me control the structures I generate in a strongly top down manner and it also acted as a binding force in my listening experience, as I tried to interpret different materials using spatiality and sound object concepts.

Instead, no single sound object representation seems to be versatile enough. Another direction can be taken – methods of generating representations are explored instead of focusing on manipulation of existing somg;e model. A variety of representations is generated by a guided process, but without focusing on the internal logics of the representation or its equivalent structural parts in the imagined sound object. The process of generating different representations can become the focus of composition, relating sounds geneologically and traversing geneological trees.

References

Barlow, H. (1972). *Dark and Light Adaptation: Psychophysics*. New York: Springer-Verlag.

Blackburn, M. (2009). *Composing from spectromorphological vocabulary: proposed application, pedagogy and metadata*. 1st ed. [ebook] Available at: http://www.ems-network.org/ems09/papers/blackburn.pdf [Accessed 13 Jan. 2016].

Chion, M. and Gorbman, C. (1994). *Audio-vision*. New York: Columbia University Press.

Cox, C. and Warner, D. (2004). *Audio culture*. New York: Continuum.

Di Scipio, A. (1994). Micro-time sonic design and timbre formation. *Contemporary Music Review*, 10(2), pp.135-148.

Gibson, J. (1933). Adaptation, after-effect and contrast in the perception of curved lines. *Journal of Experimental Psychology*, 16(1), pp.1-31.

Gunnarson, B. (2012). *Processes and Potentials: Composing through objects, networks and interactions*. Master Thesis. Institute of Sonology, The Hague.

Kane, B. (2014). *Sound unseen*. New York: Oxford University Press USA, pp.15-17.

Kennedy, M. and Bourne, J. (1994). *The Oxford Dictionary of Music*. 2nd ed. Oxford: Oxford University Press.

Koenig, G. (1968). *Remarks on Compositional Theory*. 1st ed. [ebook] Available at: http://www.koenigproject.nl/Composition_Theory.pdf [Accessed 8 Jan. 2016].

Ladd, G. (1894). Direct control of the retinal field. *Psychological Review*, 1(4), pp.351-355.

Latham, A. (2002). *The Oxford companion to music*. Oxford: Oxford University Press.

McGurk, H. and Macdonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264(5588), pp.746-748.

Mordvintsev, A. (2016). DeepDream - a code example for visualizing Neural Networks. [Blog] *Google Research Blog*. Available at: http://googleresearch.blogspot.nl/2015/07/deepdream-code-example-for-visualizing.html [Accessed 27 Nov. 2015].

Neuhoff, J. (2004). *Ecological psychoacoustics*. Amsterdam: Elsevier Academic Press.

Noë̀ˆ, A. (2004). *Action in perception*. Cambridge, Mass.: MIT Press.

Pareidolia. (2012). In: *Collins English Dictionary - Complete & Unabridged 2012 Digital Edition*, 10th ed. HarperCollins Publishers.

Russell, M. (2006). *Husserl*. London: Continuum.

Smalley, D. (2007). Space-form and the acousmatic image. *Organised Sound*, 12(01), p.35.

Smalley, D. (2014). *Spatiality in acousmatic music (lecture)*. [video] Available at: https://www.youtube.com/watch?v=_G68Q4gkOMc [Accessed 15 Jan. 2016].

Stockhausen, K. (1972). *Moment-Forming and Integration (MOMENTE)*. [online] YouTube. Available at: https://www.youtube.com/watch?v=MFvynwLkGco [Accessed 11 Dec. 2015].

The Java Tutorials. (2015). *What Is an Object? (The Javaâ„¢ Tutorials)*. [online] Available at: https://docs.oracle.com/javase/tutorial/java/concepts/object.html [Accessed 6 Nov. 2015].

Wikimedia, (2016). *Standard photograph before and after undergoing partial DeepDream processing*. [image] Available at: https://commons.wikimedia.org/wiki/File:DeepDreamingProcess.jpg [Accessed 21 Feb. 2016].