Embodied Sonification

Stephen Roddy

Thesis submitted for the degree of Doctor of Philosophy

University of Dublin, Trinity College.

2015

This research was partly funded by a

Trinity College Dublin Studentship Award.

Declaration

I declare that this thesis has not been submitted as an exercise for a degree at this or any other university and it is entirely my own work.

I agree to deposit this thesis in the University's open access institutional repository or allow the library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgement.

b) In the case of a thesis for which the work has been carried out jointly; the declaration must have a statement that it includes the unpublished and/or published work of others; duly acknowledged in the text wherever included.

c) Open access electronic theses are freely available over the World Wide Web for users to read, copy, download, and distribute subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgement. Withheld access will be implemented, as applicable, according to the terms detailed in 1.34.15.

d) A digital repository is an online, searchable, web-accessible database containing works of research deposited by scholars, the purpose of which is both increased access to scholarship and long-term preservation.

Signed:

Date:

Acknowledgements

First and foremost I'd like to thank Mary for her unwavering love and support throughout this process. This thesis is dedicated to you. Thanks to Dermot for giving me this opportunity and going above and beyond in his guidance and support. Thanks to Mum and Dad for your love support and encouragement. Thank you Sarah, James and Laura for being Sarah, James and Laura, always being there and for the help and support throughout this process. Thanks to Brian Bridges, Alan Dormer, Kelly and Danny Goddard, Lucie Karlachová, Dominik Štorek, David Poirier-Quinot, Tony Steffert, Tony Doyle and all the friends and colleagues who helped along the way. Thanks to Trinity College School of Engineering for providing me with the Studentship that made this research possible. Thanks to the members of the International Community for Auditory Display who welcomed my research and generously offered their help and guidance at Łódź, Graz and beyond. Thanks to everyone at the Irish Sound Science and Technology Convocation who have provided an invaluable forum that allowed me to get to grips with my own work and interests and to get some small insight into the wider field. Thanks to Trinity Trust Travel grant for giving me the opportunity to attend two ICAD conferences.

Table of Contents

Chapter 3 **Embodied Sonic Dimensions – Vocal Gestures** **Pages**

Chapter 4 Embodied Sonic Complexes – Embodied Soundscape Sonification Framework Pages

Chapter 5 Embodied Sonic Metaphor - The Temporo-spatial Motion Framework. Page

Chapter 6 Summary and Discussion

Appendices

List of Figures and Tables

Summary

Sonification is the communication of data using sound. This thesis is concerned with meaning-making in sonification. It examines how meaning emerges during sonification listening through the lens of embodied cognitive science. It suggests that approaches to sonification which exploit the embodied nature of meaning-making can leverage aesthetic dimensions of sound which allow for more effective communication. This thesis employs a mixed methods research approach to empirically evaluate dimensions of sound which are traditionally categorised as aesthetic. Hypotheses about the embodied nature of meaning-making in sonification listening are developed through a reading of the literature, through exploration with prototyping platforms, and through exploratory data-driven composition. These hypotheses are then submitted to empirical listener evaluations, from which conclusions about embodied meaning-making in sonification listening can be drawn.

Chapter 1 considers sonification in terms of two models of meaning-making: the computationalist model, which reduces meaning-making to information processing, and the embodied model, which treats embodied experience as the basis of all meaning. Aesthetic approaches, scientific approaches and leading sonification techniques are considered in reference to these models and it is argued that the embodied model is of more use to the development of communicatively effective sonifications than the popular computationalist model, which cannot account for the aesthetic dimensions of sound.

The Embodied Sonification Listening Model (ESLM) is introduced in Chapter 2 to offer an embodied perspective on meaning-making during sonification listening. Chapter 2 also introduces and defines the Embodied Sonic Dimension and the Embodied Sonic Complex as novel conceptual measurement schemes for working with the aesthetic and embodied dimensions of sound in a sonification context. An empirical evaluation, which describes how a listener’s embodied schematic knowledge influences their interpretation of a sonification, is also presented in this chapter. The ELSM, embodied sonic dimensions and embodied sonic complex represent useful theoretical tools for guiding the design of sonifications, which can exploit the embodied nature of meaning-making to effectively communicate data to a listener.

Chapter 3 investigates the communicative effectiveness of some of the embodied sonic dimensions afforded by vocal gestures. It presents a large number of evaluations, the results of which highlight useful synthesis strategies for modelling and controlling these

dimensions in a sonification context. The embodied sonic dimensions explored in Chapter 3 represent a means for intuitively communicating data to a listener on the basis of their previous embodied experiences.

Chapter 4 investigates the communicative effectiveness of some of the embodied sonic complexes which environmental soundscapes can offer to sonification. The research described in this chapter was intended to develop an embodied soundscape sonification framework. This was achieved through the development of a number of candidate frameworks *via* exploratory research practices, which were then submitted to rigorous empirical evaluations intended to determine the most effective framework. The embodied soundscape sonification framework presented in Chapter 4 represents an effectively communicative sonification technique, suggested by empirical evaluation to be more effective than pitch-mapping sonification in certain circumstances.

Chapter 5 explored the conceptual metaphorical underpinnings of sonification listening as described by the ESLM. It employed a similar research method to Chapter 4 to develop the "Temporo-Spatial Motion Framework" which exploits metaphorical mappings between the time and space domains to add a sense of temporal context to time-series sonification. This framework was then further refined through a process of empirical listener evaluation. The temporo-spatial motion framework represents a solution to the problem of representing temporal context in time-series sonification.

Chapter 6 summarises the research described in this thesis and a presents a discussion of its relevance and contribution to the wider field. It argues that the thesis has presented compelling evidence for an embodied approach to sonification.

Chapter 1: Embodied Sonification

1.1 Introduction and Motivations

Sonification is the communication of data using non-speech sound. It is useful for communicating the kinds of data, and creating the kinds of meanings for a listener, that cannot be easily communicated by visual means. Sonification has traditionally aimed to make use of the frequency, amplitude and spatial resolutions offered by auditory perception. The research described in this thesis explores the capacity of sound to structure meaningful experiences for a listener, in order to create communicatively effective sonifications. Sound can be used to make data meaningful by relating that data to the familiar, everyday, embodied experiences that listeners share in common. This suggests that sonification might provide a medium to communicate a broad spectrum of meanings to a large set of listeners across barriers of language, nationality and even culture, in an intuitive[1] manner. This potential suggests an embodied approach to sonification might represent a valuable and worthwhile research interest. This thesis aims to contribute to the establishment of sonification as a valid technique for understanding, communicating and interacting with data. Barrass (2012) argues that the current aesthetic turn in the field may drive the adoption of sonification as a mass cultural medium. Rimland *et al.* (2015) discuss how as societies become increasingly networked and these networks grow in size and complexity sonification will be needed as a means of making sense of the complex data flows that can no longer be effectively understood by visual means alone. It is here argued that in the future sonifications will be increasingly delivered to large international audiences of non-specific listeners across distributed networks defined by smart devices and smart sensors, traditional desktop computers, and the emerging internet of things. Sonification has the potential to become a popular method by which the invisible data flows of smart cities and environments are revealed and by which the digital environment is re-embodied in the physical environment. Listeners could turn to sonification for enjoyment, aesthetic appreciation and to learn about the data flows that define the world around them.

---

[1] The term intuitive is used in this thesis in the sense intended by Naumann *et al.* (Naumann *et al.*, 2007): "A technical system is, in the context of a certain task, intuitively usable while the particular user is able to interact effectively, not consciously using previous knowledge."

However, the field of sonification is still relatively new and is continually in development. In order to achieve this vision, much research is needed into the nature of sonification listening. This thesis makes some initial progress towards this goal by focusing on meaning-making in sonification and exploiting the embodied nature of auditory cognition to make data meaningful to a general audience through sound.

Embodied cognition is a research programme in cognitive science that describes how novel, bodily-mediated patterns of experience become meaningful for a subject when related to previous bodily-mediated patterns experiences. This thesis describes how a listener's embodied cognitive faculties, those cognitive faculties available to a listener by the nature of their physical bodies, can be exploited to create communicatively effective sonifications. It describes sonification techniques that are grounded in empirical research and exploit the kinds of meaning that are traditionally classified under the heading of aesthetics. The study of meaning-making has historically focused on linguistic meaning and the semantics, syntax and pragmatics of traditional semiotics (see Johnson, 2010). As such it has generally excluded any comprehensive account of meaning-making in aesthetic experience. The embodied turn in cognitive science has opened the door to an empirical and thoroughly objective account of meaning-making and aesthetics, and has shown that both are structured and underpinned by the common apparatus of embodied of cognition (Johnson, 2010). As such this thesis is guided by the following thesis statement:

An embodied cognitive approach to meaning-making and aesthetic practice can support the design of communicatively effective sonifications.

Chapter 1 explores approaches to meaning-making in sonification research and practice. It considers two models of meaning-making before discussing leading techniques in the area in relation to these models. Scientific and aesthetic approaches to sonification are also considered in relation to these models, before the presentation of a concluding discussion.

Chapter 2 introduces the Embodied Sonification Listening Model (ESLM), the Embodied Sonic Dimension and the Embodied Sonic Complex. It also presents an empirical evaluation that sheds some light on the role of embodied knowledge in sonification listening as described by the model.

Chapter 3 investigates the concept of the embodied sonic dimension. It also presents a set of empirical evaluations aimed to determine communicatively effective embodied dimensions, within the domain of vocal gestures that could be used for sonification.

Chapter 4 investigates the concept of the embodied sonic complex. It also presents the embodied soundscape sonification framework, which exploits the embodied nature of meaning-making in sonification listening, as described in the ESLM, to create communicatively effective soundscape sonifications.

Chapter 5 presents the Temporo-Spatial Motion Framework which is a framework for communicating temporal context to a listener in the sonification of time-series data.

Chapter 6 offers a summary of the research presented in this thesis and a discussion of its relevance and contribution to the wider field.

1.2 The Promise of Embodied Sonification

Sonification research is currently faced with a number of challenges. One challenge is the development of sonification as an empirically valid technique (see Degara *et al.,* 2013) that can be used to analyse, represent and communicate scientific data (Hermann, 2008). A second challenge is the question of how to capitalise on the aesthetic potential of sound to create more effective sonifications (Barrass, 2012; Barrass and Vickers, 2011; Serafin *et al.* 2011; Vickers and Hogg, 2006; Vickers, 2005). The literature suggests that both of these aims may be achievable within an embodied cognition framework. Lakoff and Johnson (1999) argue that scientific knowledge and research are made possible by the embodied nature of meaning-making, the process by which a person assigns meaning to their experience, and that an approach to scientific research that recognises the embodied nature of cognition is more empirically rigorous than the disembodied alternative. Johnson (2008; 2010) argues that aesthetic experience is also made possible by the embodied nature of cognition, and that a person’s meaning-making capacities are never as fully engaged as when in the comprehension of aesthetic dimensions. This suggests that embodied cognition may provide a framework in which to explore the untapped aesthetic and meaning-making potentials of sound in a sonification context. Embodied cognition provides a unifying paradigm in which sonification research and practice can exploit the aesthetic dimensions of sound to create meaning for a listener, while also being held to rigorous empirical standards.

Some more challenges are the development of a comprehensive account of the cognitive processes at work during sonification listening (Vickers 2012; Neuhoff, 2011;

Gossman, 2010; Worrall, 2009; Neuhoff and Heller, 2005; Walker and Kramer, 2004) and the need for a cohesive and unified theoretical account of sonification, a relatively new interdisciplinary field that is in a continual process of self-definition (Walker and Nees, 2011). Embodied cognition may provide a framework in which to explore the cognitive processes that underlie meaning-making in sonification listening. Johnson (1987; 2007) holds meaning-making to be the primary function of embodied cognition, a view which is echoed across the field (Varela *et al.*, 1991; Núñez and Freeman, 1999). A large portion of the literature in the area focuses on meaning-making, and describes cognition as a meaning-making activity as a result. This could be advantageous for sonification research and practice as it offers a model of cognition that is centred on what is arguably the core task of sonification: meaning-making, i.e., making data meaningful through sound.

Embodied cognition might offer an appropriate framework in which to unite the aesthetic and scientific aspirations of sonification research and practice through a shared cognitive model of meaning-making. It may also provide a context within which to develop a comprehensive and unified theoretical framework for sonification that accounts for the cognitive, aesthetic and scientific concerns of the field simultaneously.

1.3 Defining Sonification

There have been a number of different definitions offered for Sonification within the field since its inception. The Sonification Handbook broadly defines auditory display as any display that uses sound to communicate information, and sonification as a subset of auditory display that represents information by mapping data to non-speech audio (Walker and Nees, 2011). The Sonification Report (Kramer *et al.,* 1999) defines sonification as "the transformation of data relations into perceived relations in an acoustic signal for the purposes of facilitating communication or interpretation". Hermann (2008) offered an intensive four-point definition of sonification that he later simplified to "the data-dependent generation of sound, if the transformation is systematic, objective and reproducible, so that it can be used as a scientific method" (see Hunt and Hermann, 2011). Barrass offers a design-thinking inspired definition of sonification as "a mapping of information to perceptual relations in the acoustic domain to meet the information requirements of an information processing activity" (Barrass, 1997, p. 29-30). Worrall (2009) sees sonification as "the acoustic representation of data for relational interpretation by listeners, for the purpose of increasing their knowledge of the source from which the data was acquired". Scaletti defines sonification as "a mapping of numerically represented relations in some domain under study to relations in an acoustic domain for the purposes of interpreting, understanding, or communicating relations in the domain under study" (Scaletti, 1994, p. 224).

The definitions of Walker and Ness, and Kramer are instrumental in marking out the basic process involved in sonification practice and the larger scope of the sonification research programme. The rigorously empirical spirit represented in Hermann's definition is of critical importance to the development of sonification research and practice. Barrass' focus on perceptual relations, Scaletti's focus on understanding and Worrall's focus on the source of the data are pertinent to the design of communicatively effective sonifications that are meaningful to a listener. While these definitions have notable strengths, they do not expressly focus on meaning-making. Taking these points into account, this thesis proposes a definition of sonification that recognises the importance of embodied meaning-making:

Sonification is the systematic data driven generation of non-speech sound in order to communicate information about a data source to an embodied listener, who is tasked with perceiving the appropriate meaning(s) within, and/or assigning the appropriate meaning(s) to, that sound.

This definition retains the empirical spirit of Hermann's definition, Worrall's distinction between the data set and data source and Scaletti's focus on understanding discussed previously. It also asserts the importance of embodiment to sonification listening and highlights the listener's role in assigning meaning to a sonification. This definition represents a conscious attempt to place meaning-making as a key concern in sonification research and practice. With sonification thus defined the following sections of this chapter will focus on aesthetics and meaning-making. This chapter divides meaning-making into two broad categories, computational and embodied, describing how both have come about and how they assert themselves in some key sonification techniques and in the scientific and aesthetic aspects of sonification research and practice.

1.4 Computational Meaning-Making

Meaning-making is the cognitive process by which people create, perceive and assigns meaning in and to their world. There are numerous theories of how this cognitive process operates and of the faculties that are involved. The culturally and scientifically dominant paradigm for meaning making was given its strongest synthesis by Hillary Putnam (1967) and is referred to as the Classical Computational Theory of Mind (CCTM or computationalism). Computationalism argues that the human mind, as distinct from the brain, operates in the same way as a computer and cognitive tasks, such as decision-making and meaning-making, are purely computational processes, in which computation is defined by the rule based manipulation of symbols (Varela *et al.*, 1991). In this view the meaning of a symbol or perceptual input is computed by its mental manipulation in accordance with a specific set of rules.

It is often argued that the foundations of computational meaning-making are provided by mind-body dualism. This is the belief that human beings are composed of two independent parts, a physical body and the mind that inhabits it. Some commentators have pointed to Descartes' (1647) radical skepticism, as summed up in his famous "*Cogito ergo sum*", as the origin for the widespread adoption of mind-body dualism (Ryle, 1949; Damasio, 2008). Others have argued against this view claiming that radical skepticism does not directly equate to mind-body dualism (Baker and Morris, 1996). Regardless, mind-body dualism has been a thematic focus for discussion in the Western philosophical tradition (Watson, 2007).

Loeb (1981) argues that between the writings of Descartes and Hume there exists a standard theory of philosophy that describes how the Continental Rationalists, Spinoza and Leibniz, addressed themselves in turn to Descartes' mind-body dualism before the British Empiricists, Locke, Berkley and Hume, respectively advanced and finally completed the project. Loeb (1981) argues that this standard theory, although widely accepted is highly inaccurate. Watson (1993) criticises this shadow history as one of many stories about the history of philosophy that philosophers know is false, but accept as true.

Whatever the actual sequence of events, Todes (2001) argues that one majorly negative outcome of Western philosophical thinking from Descartes through Leibniz and Hume, to Kant was the propagation of mind-body dualism. Though Todes would attempt to correct this with his argument that "[t]he human body is the material subject of this world" (2001, p. 88), the distinction was already deeply entrenched in both the scientific and the common-sense view of reality in the western world. The classic philosophical example of this thinking is to be found in logical positivism. Logical positivism grew out of the Moritz Schlick-led group of 1920s thinkers called the Vienna Circle, which attempted to formalise a universal symbolic logic that could underpin and unite all areas of empirical scientific pursuit. Logical positivism drew upon a 1928 synthesis by Rudolf

Carnap (Carnap, 1928) of Wittgenstein's (1921) and Whitehead and Russell's (1912) breakthroughs in the logic of language and mathematics, respectively, to declare in its verification principle that any question that cannot be answered by logical or empirical means is meaningless. This is a profoundly limited and disembodied view that imposes severe restrictions on the scope and definition of meaning and which entirely overlooks the role of embodied human experience in meaning-making. This view would exert a strong influence over the way meaning and meaning-making are approached and conceptualised in philosophy, computer science, computer music, psychoacoustics, music, cognitive science and psychology (Lakoff and Johnson, 1980; 1999), all of which are fields involved in sonification (Hermann *et al.*, 2011).

Gardner (1985) describes how computationalism became a driving force during the early days of cognitive science. Cognitive science, the study of the mind through the multidisciplinary application of the scientific method, began to take shape in the late 1940s during a period of growing dissatisfaction with behaviourism, which was the leading research programme in psychology at the time. Behaviourism focused solely on the study of publicly observable behaviours and stimuli. It rejected the study of subjective and privately observed elements of cognition such as concepts, thought processes, emotions, imagination and even the very idea of a mind. This was a reaction against the failures of structuralism, an earlier research programme that sought to understand the human mind through the systematic analysis of subjective experience, an approach that proved hard to implement, and produced little empirically verifiable or reproducible evidence (Vermersch, 2009). By 1948 a growing number of researchers within both psychology and the wider scientific community were concerned that the overly restrictive scope of behaviourism was preventing any meaningful study of the human mind and began to search for an alternative (Gardner, 1985).

Researchers interested in the mind looked to the emerging field of computer science to provide a new model of cognition. Computer science underwent a period of rapid development due to progress driven by the nuclear arms race that took place towards the end of World War Two and scientists in the area had begun to draw parallels between the computer and the human mind. In 1936 Turing (1936) introduced the universal Turing machine, a hypothetical device that could simulate the logic of any algorithm using four simple rules. That same year the Church-Turing thesis lent formal definition to the concept of the algorithm (Church, 1936, Turing 1936). This laid some of the initial groundwork for the coming developments in computer science. In 1937 Claude Shannon

(1937) showed that Boolean logic could be used to represent states in electromechanical relay switches and that those switches could be used to solve logical problems. On this basis he suggested that relay switch circuits might also be used to model cognitive processes. In 1943 Warren McCulloch and Walter Pitts (McCulloch and Pitts, 1943) presented a theorem which showed how networks of neurons, could be modelled in terms of propositional logic. They concluded that all sensation and mental activity can be described in terms of binary logic and that the universal Turing machine would be capable of running these logical processes. In 1948, mathematician Norbert Wiener (Wiener, 1948) introduced the field of cybernetics defining it as the scientific study of control and communication in the animal and the machine, he drew comparisons between self-regulation and self-correction across electronic, mechanical and biological systems. Wiener argued that machines, which exhibited feedback, could be described as striving towards goals because they modified their own behaviour to achieve an objective. The Turing Test (Turing, 1950) proposed that if a user could not distinguish the responses of a machine from those of a human, then that machine could be said to be capable of thought. Breakthroughs in computer science during this period seemed to be making the subjective and private cognitive processes ignored by the behaviourists publicly observable and objectively verifiable. By the 1950s there was a growing sense that advances in computer science were revealing exciting new insights about the human mind. The first generation of cognitive science was heavily influenced by these developments and adopted a computational model of the mind. It rendered perceptual and mental content as symbols on a layer of mental representations that existed between sensory inputs and behavioural outputs. Thought was conceptualised as the computation or rule based processing of these symbols (Lakoff and Johnson, 1999).

The computational model of the mind had a number of shortcomings that negatively impacted the field of artificial intelligence (A.I.). In 1973 the British Science Research Council approached Professor Sir James Lighthill, an aeroacoustics scientist, to undertake an impartial state of the art review of the A.I. field. Lighthill concluded that A.I. had made little progress since the publication of Turing’s Intelligent Machines article (Turing, 1947) 25 years earlier (Lighthill, 1973). Lighthill argued that A.I. algorithms were far too intractable to effectively deal with the complexity of real world conditions and concluded that computation might never be capable of simulating the self-organising capabilities of the nervous system. In the wake of the Lighthill report the British government pulled funding for all but three university A.I. programmes while in the U.S. DARPA (the

Defense Advanced Research Projects Agency) channeled its own A.I. funding into mission-oriented research.

Alongside the failures in A.I. a number of critics voiced their concerns about the severe limitations of the computationalist theory of mind. Ryle (1949) argued that computation could not simulate intelligence. He criticised the definition of intelligent acts as those driven by the prior computation of logical rules because any prior computation, must also be driven by a prior computation in order to be intelligent. This would result in an infinite regress. Humans experience no such regress when acting intelligently and so computation alone could not account for intelligence. Dreyfus (1965) argued that symbolically mediated cognitive processes require a context of tacit, informal background knowledge to render them meaningful. A large portion of human knowledge, for example domain specific expertise, is tacit and informal and so cannot be represented symbolically thus computation alone cannot account for knowledge with a tacit component (for further information see Dreyfus, 1992). Searle's Chinese Room problem (Searle, 1980) also highlights the limitations of a computationalist approach to meaning-making. It shows that rule-based computation may be sufficient to pass the Turing test, but that computation alone can never account for how it is that symbols are assigned their meanings. Harnad (1990) formalises Searle's problem of how symbols become meaningful as the symbol-grounding problem. He further argues that human experience is full of meaningful symbols for which computationalism cannot account and so it is not a sufficient theory of mind or meaning-making.

The symbol-grounding problem, and the shortcomings of the computationalist theory of mind that it exposes, are relevant to sonification. Sonification is concerned with representing data sources using sounds, which can be conceptualised as symbols. In order for a sonification to be communicatively effective, that is to achieve the intentions of the designer in the mind of the listener, the sounds used must encourage the listener to assign the correct meanings. Considering only the computational aspects of cognition overlooks meaning-making and the processes by which the sonic symbols in a sonification become meaningful. Researchers have also pointed out that the computational model of mind excludes a thorough explanation of the meaning-making processes involved in aesthetic experience (Lakoff and Johnson, 1999; Johnson, 2010), a domain of human experience and scholarship that is crucial to the study of sonification (Vickers, 2012). Worrall (2010) also argues that computationalism is to blame for the mapping problem encountered in auditory display research, a point explored in greater detail elsewhere in this thesis. The following section will explore an alternative to computationalism in the form of embodied meaning-making and will also consider the solution that it provides to the symbol-grounding problem.

1.5 Embodied Meaning Making

> […] the nature of our bodies and our physical and cultural environment imposes a structure on our experience, in terms of natural dimensions […] Recurrent experience leads to the formation of categories, which are experiential gestalts with those natural dimensions. Such gestalts define coherence in our experience. We understand our experience directly when we see it as being structured coherently in terms of gestalts that have emerged directly from interaction with and in our environment. We understand experience metaphorically when we use a gestalt from one domain of experience to structure experience in another domain.
>
> -Lakoff and Johnson, 1980

Embodied cognition is a research strand within cognitive science that studies the relationship between the body and mind with a specific focus on how the physical and perceptual dimensions of the human body shape and define the cognitive dimensions of the human mind. The principle of EC are reflected in the above quote by Lakoff and Johnson, two of the founding figures of what has become the second generation of cognitive science. EC arose to prominence in the late 20th century as discontent grew with the successive failures of computationalism discussed earlier and its inability to offer adequate explanations for emotion, culture and aesthetic experience. EC would have a large impact in some of the disciplines related to sonification research and practice, e.g. computer science, artificial intelligence and human computer interaction (Brooks, 2003; Dourish, 2004; Imaz and Benyon, 2007), computer music (Leman, 2008; Klemmer *et al.*, 2006), cognitive sciences (Varela *et al.*, 1991), visual perception (Noë, 2009), aesthetics (Johnson, 2013), music (Godøy, 2006; Zbikowski, 2005; Brower, 2000; Larson, 2010; Cox, 2001), linguistics and philosophy (Lakoff and Johnson, 1999).

This thesis adopts the embodied view of meaning-making, presented in the work of Mark Johnson (Johnson, 1987; 2008; 2010; 2013; Johnson and Rohrer, 2007) and Francisco Varela (Maturana and Varela, 1987; Varela *et al.,* 1991). Both Johnson and Varela argue that meaning emerges in the coupling between an organism and its environment, as opposed to being an entirely mental phenomenon, and so is mediated and shaped by bodily interactions with the environment.[2] Johnson presents a comprehensive theory that accounts for symbolic, linguistic and conceptual meaning, and also the kinds of meaning associated with emotions, felt qualities of experience, and aesthetic experiences of art and music. He also isolates embodied schemata, also referred to as image schemata, and conceptual metaphor as embodied cognitive faculties crucial to meaning-making. These faculties will be discussed in more detail shortly.

Johnson and Rohrer (2007) claim that the "evolutionary embeddedness of the organism within its changing environments, and the development of thought in response to such changes, ties mind inextricably to body and environment". This builds on the work of Varela *et al.* (1991) who present the similar argument that meaning emerges in the reciprocal relationship, termed structural coupling, between organism and environment, as organisms evolve bodily mediated minds to aid in effectively asserting themselves in their environments. In this view, any cohesive account of meaning-making must take the role of the body into account, because meaning-making is mediated by the human body and emerges in the interaction between that body and its environments. Johnson (2013) refines his definition of meaning making further in the argument that the meaning of an event, object or symbol is defined in relation to any bodily mediated past, present and possible future experiences it offers a subject.

Critical to this theory of meaning-making is the philosophy of experientialism or experiential realism, which was first discussed by Lakoff and Johnson (1980). Experientialism is the idea that experience is the source of all meaning and that no meaning can exist in a form that is abstracted from experience. Experientialism provides an alternative to the objectivist position that reality exists independently of the human mind and that a concept or belief is meaningful to the degree that it matches its real world counter part. It also provides an alternative to the subjectivist view that reality is a purely mental phenomenon and so a person can assign any meaning they choose to the contents of experience. These ideas relate closely to Husserl's (1913) lebenswelt, Heidegger's (1927) dasein, Dewey's (1934) lived experience and Merleau-Ponty's (Ponty, 1968) chiasm.

The embodied approach argues that symbols, sonic or otherwise, become meaningful when they are associated with embodied experiences with reference to which

---

[2] This approach differs from ecological psychoacoustics in its focus on the role of the body as the conduit between mind and world. This difference is described in greater detail in Chapter 4.

they can be understood. Take for example the sound of an approaching rainstorm. The meaning the listener assigns to the sound will depend on the network of previous experiences and projected future experiences they associate with that sound and similar sounds that they have encountered in the course of their everyday lives. There may be a large degree of variance between the meanings assigned to a sound by different individuals in different contexts but there will also be a degree of commonality on the basis of shared embodiment. This is because the physical and perceptual dimensions of the human body provide a collective context, in terms of commonly shared patterns of experience (Lakoff, 2012). For example, a farmer might associate the sound of an approaching rainstorm with a good source of crop irrigation and a holiday maker might associate it with an unwelcome interruption to a holiday but both are likely to associate it with the experiences of being rained on and getting wet. Embodied cognition researchers argue that shared patterns of experience, termed embodied schemata, provide the basis by which people assign meanings to the perceptual and conceptual content of their worlds (Johnson, 1987; Lakoff, 1987; Lakoff and Johnson, 1999; Johnson, 2008). The relative invariance of embodied schemata from person to person, allows people to assign meaning to their worlds in a in a similar and systematic manner. These embodied schemata and the ways in which they are used to assign meaning are integral to the embodied account of meaning-making.

1.6 Embodied Meaning-Making Faculties

Embodied cognitive science has proposed and modelled a number of cognitive meaning-making faculties thought to emerge in the shared relationship between similarly embodied organisms and the environment. Embodied schemata, first discussed by Johnson (1987) and Lakoff (1987), are commonly shared fundamental gestalt patterns of embodied experience that provide people with a common basis for organising their experience, meaning-making and reasoning. Johnson and Rohrer (Johnson and Rohrer, 2007) further describe embodied schemata as “recurrent patterns of bodily experience, which preserve the topological structure of the perceptual whole”. Embodied schemata impose an elaborate and detailed structure on the chaos of raw experience independent of and prior to the processes of conceptualisation and language. Each schema has its own internal logical syntax, which provides a basis for reasoning and inference. For example Johnson’s source-path-goal schema (Johnson, 1987) describes the pattern shared by experiences in which a trajector, an entity that follows a trajectory, departs from a source and moves along a path towards an ultimate goal. In the internal logic of the source-path-goal schema the source always precedes the goal and in order to get to a goal a path must be traversed. From this it can be reasoned that if a trajector is on the path then it has departed the source and is not yet at the goal and if a trajector is at the goal it can be inferred that it has departed from the source and traversed the path. These logical syntaxes organise experience into meaningful relations and can be used to lend structure to unfamiliar conceptual domains. In recent years they have be used in the design of intuitive computer interfaces (Imaz and Benyon, 2007; Hurtienne and Blessing, 2007).

There is a wealth of empirical evidence to support the claim that certain embodied schemata are common to large populations of people at a pre-linguistic level (Hampe and Grady, 2005; Johnson, 2013; Lakoff and Johnson, 1999). On the neural level they are realised as activation patterns in and between topologic neural maps and they are used to link sensorimotor experience to thinking and emotionality. The cross-domain mapping of embodied schemata across physical, perceptual and conceptual domains may provide a crucial mechanism by which sonic patterns get their meaning.

The concept of mapping appears repeatedly across the embodied cognition literature (Lakoff and Johnson, 1999; Fauconnier and Turner, 2002). Mappings associate content from one mental space[3], or domain of embodied human experience with content in another. It is the basic process by which perceptual and conceptual symbols are assigned meaning. For example in the concept of a red herring the concept red is mapped from the domain of colour onto the concept of a herring from the domain of fish and for a person who is aware of the cultural connotations of the term, the concept of a decoy is also mapped onto the red herring.

Conceptual metaphors are a specific type of cross-domain mapping where embodied schemata from familiar areas of experience, termed source domains, are mapped onto unfamiliar target domains that would otherwise be meaningless or unknowable, in order to make them meaningful (for further details see Lakoff and Johnson, 1980). The classic example of a conceptual metaphor is the LOVE IS A JOURNEY[4] metaphor in which the source-path-goal schema underlying a subject's experiences of journeying is mapped to lend familiar structure to the abstract concept of love. This allows love to be

---

[3] A mental space is a broad domain of related embodied schematic knowledge.

[4] Capitalisation is used in cognitive science literature to denote a conceptual metaphor.

conceptualised as a journey with a beginning, middle and end where the lovers are travellers on a common path along which they may encounter difficulties and perils. In a conceptual metaphor, all source domains are ultimately grounded in the embodied schemata of sensorimotor experience (Lakoff and Johnson, 1980).

The concept of the conceptual integration network or blend was introduced by Fauconnier and Turner (2002) to describe how new structures of meaning can be created from basic embodied schemata during acts of creative and artistic thinking. A blend cross-maps conceptual content and embodied schemata from one mental space to another, thereby creating entirely new mental content that represents a blend of the content in the input spaces (For a more detailed analysis see Fauconnier and Turner, 2002). For example the mythical concepts of the Pegasus and Centaur have been described as blends between the concepts of a bird and a horse and the concepts of a man and a horse respectively (Martinez *et al.*, 2012). Kendall (2014) uses blending theory to offer an embodied account meaning-making in electroacoustic music in which he extends the theory to account for the novel emotional and phenomenal qualities that emerge during electroacoustic listening as blends of familiar emotional and phenomenal qualities. The concepts of embodied schemata cross-domain mapping, conceptual metaphor and blending are central to the embodied cognition paradigm. These faculties are thought to organise both perceptual and conceptual experience into an intelligible and meaningful framework. It is argued that they are more than useful tools for interpreting and understanding the world, but are the faculties by which the experience of any intelligible world at all is made possible (Lakoff and Johnson, 1999; Fauconnier and Turner, 2002).

1.7 Neural and Sensorimotor Architecture

Neural networks in the human brain and sensorimotor system carry out the complex computations which drive these cognitive faculties but the phenomenological experience of the human mind is not characterised by this neural computation (Lakoff, 2012). Instead it is characterised by knowing and understanding; feats it can only achieve by recruiting the neural circuitry of bodily perception and action (Varela *et al.*, 1991). A central claim of embodied cognitive science is that there is no fundamental separation between mind and body and so our physical embodiment shapes thinking and conceptualisation on both the neural level and the phenomenological level as described previously. Lakoff (2012) argues that a wealth of empirical results on the embodied nature of cognition support claims that thought is physical and undertaken by neural circuitry associated with perception and action and it is the specific ways in which those circuits connect to the body and characterise embodied experience that lends meaning to the symbols of language and thought. He further argues that embodied cognitive science show that embodied schemata, conceptual metaphors and conceptual blending are driven by a process of sensorimotor mimesis at the neural level. This process of sensorimotor mimesis is referenced repeatedly across the embodied cognition literature (Lakoff and Johnson, 1999; Johnson and Rohrer, 2007; Johnson, 2008) it involves the internal mimicry of patterns of neural activity associated with gesture, perception and proprioception at an unconscious level within the nervous system. Johnson (1987) also argues that this is the process by which embodied schemata are mapped from one domain of experience to another at the neural level. This echoes Cox (2001) who argues that meaning-making in music listening is driven by a similar process of miming musical and sound producing gestures in the nervous system in order to assign meaning to patterns of sonic and musical experiences. A further discussion of the embodied neural architecture of relevance to sonification research and design has been provided by Worrall (see Worrall, 2010, 2011).

On the phenomenological level embodied schemata and cross domain mapping provide the mechanism by which perceptual patterns are assigned meaning. Meaning is assigned by associating familiar patterns of perception with the previously encountered embodied schemata they resemble generating complex networks of possible meanings (Johnson, 1987). The process by which a concept is assigned meaning leverages the same perceptual apparatus by mapping embodied schemata to conceptual domains to lend meaning to conceptual content. In acts of understanding, the meaning of a familiar domain of experience is mapped onto an unfamiliar conceptual domain through conceptual metaphor (Lakoff and Johnson, 1999). This point is especially important for the practice of sonification listening in which a listener attempts to indirectly understand a data source with reference to a configuration of novel sonic symbols. In acts of creative thinking new conceptual content is created through the process of conceptual blending (Fauconnier and Turner, 2002). The relationship of these faculties to sonification is explored in greater detail throughout the remainder of this thesis.

1.8 Embodied Symbol Grounding

The symbol grounding problem, discussed earlier, asks how meaning is assigned to a symbol. The embodied model of the mind discussed in this chapter argues that meaning is assigned to abstract symbols by associating them with patterns of embodied experience with reference to which they can be understood. The embodied schematic structural patterns and relationships evident in the symbol provide the basis for interpretation.

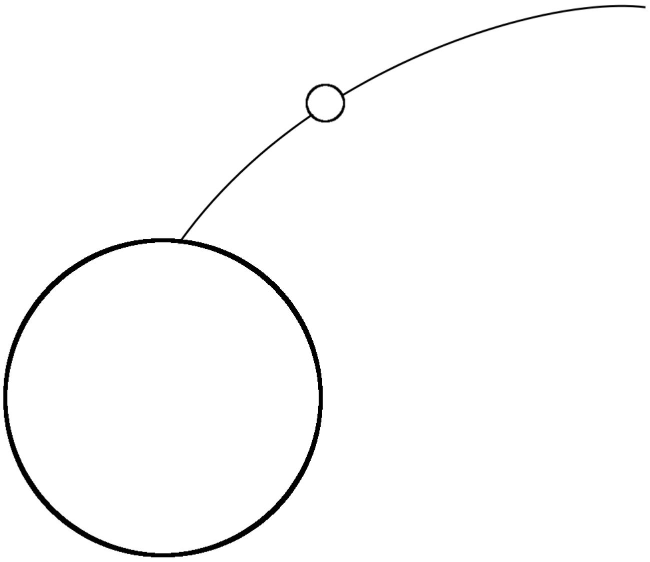


Figure 1.1. Abstract Symbol

For example, Figure 1.1 represents an abstract symbol. The term abstract is used in this thesis to describe any novel, unfamiliar symbol for which no extraneous referential context by which meaning might be assigned to that symbol is supplied. Abstract symbols can be assigned meaning on the basis of familiar patterns and features identified in the structure of the symbol. Figure 1.1 can be interpreted in terms of the patterns and logics of the object, container, big-small, up-down, left-right and source-path-goal schemata (Johnson, 1987). Both the big and small circles can be conceptualised as objects. Both circles can also be conceptualised as containers and the smaller circle as an object which is outside of the big circle's area of containment and located above and to the right of it. The symbol can also be conceptualised in terms of the source path goal schema with the line that passes through the small circle and links it to the large circle representing a path. The location of the source and the goal in this instance are not evident and so the small circle can be conceptualised in two ways, as moving either towards or away from the large circle. Through the process of conceptual metaphor many interpretations of the symbol are possible. The symbol could be framed in terms of biological cells where a small sodium ion is leaving a larger cell or in terms of astronomy where a comet is approaching the planet. This has importance for sonification as it suggests that abstract sonic symbols might be interpreted with reference to a contextual background of embodied schemata and conceptual metaphors. Multiple interpretations of the abstract symbol in Figure 1.1 are possible and as such additional information must be added to the symbol to ensure that it is interpreted in reference to the contextual background of embodied schemata and conceptual metaphors. This can be achieved by the addition of context cues which indicate as represented in the use of labels in Figure 1.2.

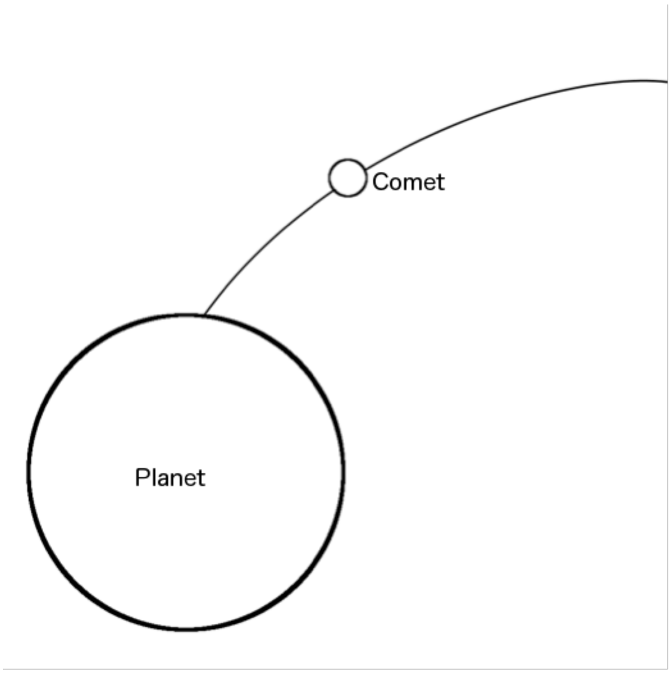


Figure 1.2. Labelled Symbol

This adds a second layer of meaning to the symbol with reference to which elements of the symbol can be interpreted. Interpreting context cues of this nature is a simple task in visual graphing practices where both the symbols and the labels of the graph endure in time and can be attended to and interpreted in relation to one another in a time independent manner. This is not the case in a sonification context where the sound is in a continual process of change and evolution. It is here argued that in sonification the structure and behaviour of the sonic symbols must clearly indicate to the listener the correct contextual background of embodied schemata and conceptual metaphors with reference to which those symbols should be interpreted. This can be achieved by leveraging the aesthetic dimensions of the symbol to indicate the correct context against which that symbol can be interpreted as illustrated in Figure 1.3.

Figure 1.3. Grounded Symbol

1.9 Embodied Aesthetics

A number of embodied cognition researchers link embodied cognitive faculties to aesthetic experience in visual and sonic mediums. Johnson (1987) argues that art is about meaning-making and that as a result aesthetic experiences are those in which the audience's cognitive meaning-making faculties are used to greater meaning-making effect than they are in standard everyday experience. He also illustrates how works by Kandinsky and the Luba people of North Shaba region of the Congo River Basin share a common embodied schematic syntax that determines the qualitative experience of these very different works. Johnson and Larson (2003) examine the embodied schematic syntax that lends The Beatles' *Something in the Way she Moves* its qualitative feel. Zbikowski (2005) examines the embodied schematic syntax, cross-domain mappings and conceptual metaphors underlying works by Mozart and Beethoven and describes how they drive the qualitative dimensions of their work. Kendall (2010; 2014) describes how embodied schemata and conceptual blending drive the aesthetic qualities across works by a number of well-received electroacoustic composers including Barry Truax and Denis Smalley. These researchers argue that aesthetic experience happens when the aesthetic syntax, of a piece is grasped and the relationship between elements of the piece and that syntax, be they points of deviation from or adherence to the syntax, are perceived. Qualitative, feeling and emotional dimensions established by the syntax are modulated by the degree to which a piece of work deviates from that syntax, a view similar to that of Meyer

(1956). This process of grasping the embodied syntax and perceiving relationships between it and elements of a piece recruits the same perceptual and neural apparatus as recruited by in the acts of knowing and understanding discussed previously (Johnson, 2008). The embodied syntax of any one work of art is determined by the multi-layered configuration of embodied schematic structure presented within the piece. Embodied schemata provide a set of aesthetic building blocks, or a meta-syntax, by which any number of unique syntaxes can be defined. Cross-domain mapping, conceptual metaphor and conceptual blending provide the methods by which these building blocks integrate to define the syntax of an individual work. Ultimately the distinction that the embodied view of meaning-making draws between normal experience and aesthetic experience is one of scale rather than kind. In the EC view described in this chapter the same meaning-making faculties drive aesthetic experience and normal experience. In aesthetic experiences these faculties are more fully engaged and generate richer networks of meaning, while in normal experiences these faculties are less engaged and generate less vivid networks of meaning.

This embodied take on aesthetics has implications for sonification. It suggests that the dimensions of sound traditionally thought of aesthetic are driven by the listeners embodied meaning-making faculties and so sonifications which effectively exploit these faculties will produce aesthetic experiences for the listener. This is important as it means that designing a sonification for the embodied mind is synonymous with designing a sonification that holds aesthetic significance for a listener. The more effectively a sonification communicates meaning to a listener, the more that the sonification listening process will become an aesthetic experience[5]. Chapters 3 and 4 explore this concept further through the composition of data-driven musical pieces in which sonification strategies designed for the embodied mind described in this chapter result in aesthetically engaging pieces.

1.10 Meaning-Making and Sonification Techniques

Having discussed two key meaning-making theories, previous sonification research and practice will now be considered from the point of view of each. The aim here is to assemble a picture of how meaning-making is approached within the field.

---

[5] Care must be taken to ensure that the correct meanings, those that represent the data source, are being communicated.

First, some sonification techniques will be examined before the field itself is considered in relation to a number of the disciplines in which sonification research and practice is pursued.

Gaver (1989) drew upon Gibson's theory of affordances to develop the auditory icon, a sonification technique that maps data to familiar everyday sounds that are often derived from the environment and so require little learning on the listener's part. Auditory icons are common in the field with heavy usage in human computer interaction (HCI) applications and are discussed in depth with a dedicated chapter in the Sonification Handbook (Brazil and Fernström, 2011). The sounding object project (Rocchesso and Fontana, 2003) expanded upon the concept of the auditory icon by introducing a physics based sound modelling approach creating hyper-real representations of acoustic events that could be easily matched with physical models and encapsulated in objects and environments. Imaz and Benyon (2007) argue that visual icons used in GUI's operate as conceptual metaphors for the tasks they represent. In the same way auditory icons and sounding objects might also operate as conceptual metaphors for the data sources that they represent. Auditory icons and sounding objects make meaning by relating data to familiar everyday sonic experiences. Brazil and Fernström (2006) present empirical evidence supporting the claim that listeners better recognise sounds which adhere to the principles of embodied cognition, auditory scene analysis (ASA) and ecological psychoacoustics simultaneously and also call for the development of auditory icons which are aesthetically engaging.

McGookin and Brewster define earcons as "short, structured musical messages, where different musical properties of sound are associated with different parameters of the data being communicated" (2011). They are also prevalent across the field of HCI where they are used to designate software events and interactions. As discussed previously, a number of researchers have demonstrated how music is driven by the listener's embodied meaning-making faculties (Brower, 2000; Zbikowski, 2005; Johnson, 2008; Larson, 2010; Kendall, 2010, 2014). Using musical content in a sonification or auditory display context might leverage the listener's embodied meaning-making faculties. Listeners might assign meaning to an earcon on the basis of its adherence to or deviation from the embodied schematic syntax it establishes as described in section 1.9.

Spearcons represent another type of auditory display technique, in which menu items or lines of written text in a graphical user interface are converted to speech (using TTS or text-to-speech technologies) and the resulting audio is sped up until it is no longer recognisable as speech. This creates novel patterns of sound that are unique to the items they represent. Listeners learn to associate the novel sonic patterns with the menu items to which they correspond. Menu items that contained similar words will share some sonic content in common thus aiding the listener in the processes of learning and identifying new spearcons. Spearcons operate differently to auditory icons and earcons. Where auditory icons get their meaning from the environments to which they metaphorically relate and earcons get their meaning from the embodied schematic syntax underlying the musical content spearcons represent an entirely new type of sound that is meaningful in reference to the menu item it is derived from.

Audification is an auditory display technique that is defined by Walker and Kramer as “the direct translation of a data waveform into sound” (Walker and Kramer, 2004). Typically, the effectiveness of an audification at communicating the data is strongly dependent on the structure of the data set, and to this end Dombois and Eckel (2011) suggest that it is most useful for highlighting subtle changes in a data set and present a set of criteria for choosing appropriate data. In audification meaning emerges when the heard sonic pattern differs from the expected pattern, indicating that there has been a similar change in the data set (see Dombois and Eckel, 2011).

Parameter mapping sonification (PMSon) maps data to auditory parameters such as pitch, amplitude, duration or timbre in order to communicate the original data to a listener (Grond and Berger, 2011). Designing a mapping strategy (the strategy by which data is mapped to sound) is both a critical and a complex task that requires numerous cognitive and physical variables to be correctly balanced (Worrall, 2013). Some of the difficulties that come along with PMSon are discussed later in this chapter. Conceptually, PMSon is a highly configurable and flexible technique that can be applied to a diverse set of sonification tasks and to represent a diverse set of data types. As a result, it is not inherently bound to either of the two broad models of meaning-making discussed in this chapter. PMSon places the mapping strategy in the hands of the sonification designer allowing the designer to decide the kind of meaning-making model they wish to adopt. This allows for the possibility of PMSons that exploit the full range of embodied meaning-making. Gaver (1993) mixed the PMSon and auditory icon techniques, achieving parametric control of auditory icons, by developing synthesis algorithms for everyday sounds. Fitch and Kramer (1994) provide an example of an effective PMSon for helping users to monitor and respond to medical complications across eight physiological variables of a digital patient. Two physiological variables are mapped to control sounds to which they resemble. Heart rate was mapped to a rhythmic thudding sound and breathing rate was mapped to a breath like sound. Atrio-ventricular dissociation and fibrillation were mapped to modulate the heart rate sounds in the same way these factors modulate a heartbeat in the real world. These mappings leveraged the users previous knowledge embodied knowledge of human anatomy. Four other mappings, body temperature to a filter applied to the heart beat sound, blood pressure to the pitch of the heart sound, brightness of the heat sound to CO2 level and pupillary reflex to a high pitched tone, were abstract and so required learning on the part of the listener. Empirical evaluation showed the auditory display system to be more effective than a visual display for helping users monitor and respond to changes in the condition of the digital patient.

Model Based Sonification (MBSon) was introduced by Hermann (Hermann and Ritter, 1999) and integrates concepts from physical modelling (see Cook, 2002) and sonic interaction design (see Franinović and Serafin, 2013). It is a general term intended to describe any Sonification technique that uses dynamic sound-making models that are parameterised during initialisation with the data and offer modes of excitation to a user. When the model is excited through user interaction it creates a sonic profile that is an expression of the original data as a function of the excitation of the model. A more in-depth discussion of the technical implementation of MBSon along with numerous examples is presented by Hermann (2002). MBSon has the potential to exploit the full range of the listeners' embodied meaning-making faculties because it allows designer to design models that best fit a users cognitive faculties.

Sonic Interaction Design (SID) is not a sonification technique, but a field of study. It merits a closer look as a sizeable portion of the work done in sonification has an interactive element. Dourish (2004) proposes 'embodied interaction' as the creation, manipulation and sharing of meaning through engaged interaction with artefacts, a view influenced by the embodied phenomenology of Mearleau-Ponty and a consideration of meaning-making in an interaction context. Many SID researchers have adopted this model of interactive meaning-making (see DeWitt and Bresin, 2007; Polotti *et al.*, 2008; Kabisch *et al.*, 2005; Rocchesso and Bresin 2007; Bovermann *et al.*, 2006; Rocchesso *et al.*, 2009; Wakkary *et al.,* 2005; Droumeva and Wakkary, 2008; Droumeva *et al.*, 2007). Antle *et al.* (2011) take embodied interaction a step further by making use of embodied schemata and conceptual metaphors to link sound and interaction supporting reasoning in an interactive sonification system. Embodied interaction addresses meaning-making in the interaction context exclusively, and does not offer an adequate description of how sound can exploit embodied meaning-making to effectively communicate data to a listener.

PMSon, the mapping of data to familiar everyday sounds as pursued in auditory icon design, and the mapping of data to musical structures as pursued in the design of earcons all show promise as useful techniques that support an embodied approach to meaning-making in sonification. These approaches are not without their problems. For example, the mapping problem discussed in Chapter 3 is an obstacle to effective use of PMSon. Even so, these techniques do have the potential to support an embodied approach to meaning-making that can exploit dimensions of sonic experience traditionally described as aesthetic. Although MBSon might also support an embodied approach, the necessity for interaction places it beyond the scope of this project because this project is focused solely on the dimensions of sonic experience involved in sonification listening and the meaning-making faculties involved. The same is true for an SID approach, while the conceptual and practical limitations of audification and spearcons leave little room for embodiment. For an additional consideration of embodied cognition as it relates to sonification research and practice see Roddy and Furlong (2013a; 2014).

1.11 Meaning-making and the Science of Sonification

Popper (1970) argues that "we approach everything in the light of a preconceived theory" and Kuhn (1962) argues that scientists rarely challenge the assumptions of the dominant paradigms into which their work falls. The meaning-making paradigms underpinning sonification research and practice are rarely examined or made available for examination in academic publications, with the exception of those where meaning-making is the explicit topic. The current and next sections aim to examine scientifically and aesthetically oriented sonification research in terms of the two models of meaning-making described earlier, in order to better understand how meaning-making is approached in sonification. Scientifically oriented sonification research is defined here as empirical research that is primarily concerned with the production of knowledge that is specifically intended to aid the design of more communicatively effective sonifications. Aesthetically oriented sonification research is defined here as any sonification research or practice that is primarily concerned with sonification as a medium for creating aesthetic experiences.

In scientific research, a large degree of emphasis if often placed on the information-processing component of sonification listening. For example, Effenberg (2004) agues that in the context of supporting motor learning and motor control tasks that drive human movement patterns; conscious and unconscious forms of information processing primarily guide sonification listening. Walker and Nees (2011) argue that the major constraining factors in sonification are perception and information-processing capabilities. Hermann and Hunt (2004) argue that learning how to use an interactive sonification is driven by information processing tasks that link physical action with sonic reaction. Nees and Walker (2011) describe listening to in-vehicle auditory display as an information-processing task (see also Nees and Walker, 2008; Fergusson *et al.,* 2011). This may be reflective of the prevalence of a computationalist model of meaning-making in the more scientifically oriented areas of sonification research and practice. As discussed earlier, the computationalist paradigm was a dominant force, and still is, in computer science, psychology and cognitive science, disciplines that Hermann *et al.* (2011) identify as critical to the development of sonification. It may be the case that sonification has assumed the some elements of the computationalist paradigm, at least in part, from these disciplines. Another discipline identified by Hermann *et al.* (2011) as critical to sonification is psychoacoustics.

Psychoacoustics studies the relationship between stimuli and their auditory perceptual responses. Established in the late 1800s by Gustav Fechner and taking inspiration from the philosophy of Spinoza and the experimental psychology of Ernst Weber it reached maturity in the 20th century with the work of Stanley Stevens (Heidelberger, 2004; Stevens, 1975). It has since become a core concern in sonification (see Carlile, 2011).

The field, as described by Walker and Kramer (2004) and Carlile (2011), would eventually grow to encompass diverse topics such as spatial hearing and Bregman's (1990) auditory scene analysis (ASA). ASA is a model of auditory perception, which describes how the auditory system organises streams of sound into perceptually meaningful patterns. It represents a pragmatic approach to auditory perception that unites elements from both the computational and embodied cognition paradigms. Bregman (1990, p. 3) argues that perception functions to generate a useful representation of external reality upon which calculations can then be performed, a prime example of a computationalist model of meaning-making (see Lakoff and Johnson, 1999). ASA also employs gestalt theory, a prime example of an embodied approach (Johnson, 1987), to auditory perception. Bregman originally conceived of ASA in the context of low-level perception without any reference to the emergence of meaning. Embodied cognition researchers argue that meaning first emerges in low-level perception in the form of embodied schemata and is later extended to higher levels (Lakoff and Johnson 1999). As such a consideration of ASA is critical to the design of intelligible auditory displays, which seek to exploit the embodied nature of auditory cognition.

Steven's psychoacoustic scaling techniques, magnitude estimation and cross-modal matching were built around his operationalist worldview, which holds that scientific concepts which cannot be measured are meaningless (Miller, 1975; Ribes-Iñesta, 2003). Psychoacoustic scaling seeks to formalise the relationships between the extents of physical stimuli and the perceptual extents they motivate. The techniques became popular listener evaluation methods in sonification research (see Walker, 2002; Walker and Lane, 2000; Smith and Walker, 2005). Being exclusively focused on quantitatively measurable extents between stimuli and percept within an operationalist paradigm, these techniques may not lend themselves well to measuring the dimensions of meaning discussed across the EC literature. Walker and Kramer (2004) argue that the classic psychoacoustic framework falls short of accounting for the complexities of hearing in real –world environments. They suggest instead that approaches, which can account for the ecological embeddedness of the listener, as described by Gibson's ecological perception (Gibson, 1978), may be favourable. This is compatible with an embodied approach to meaning-making as Varela *et al.* (1991) point out that Gibson's theory of environmental affordances, in which the environment mediates and determines the activities of its inhabitant subjects, parallels the embodied view of meaning-making where constraints imposed on the mind by bodily mediated environmental experiences define an organisms cognitive faculties.

Auditory imagery is an important facet of embodied sonic meaning-making. Auditory images are sonic experiences imagined in the absence of acoustic stimuli (Intons-Peterson, 1992). They have been shown to simulate auditory cognition in many of the brain areas associated with auditory perception (Halpern *et al.*, 2004; Hubbard, 2010), and fMRI studies have shown how thinking about sounds causes the brain to replay the same patterns of activity estimated to be present when listening to that sound (Kiefer *et al.,* 2008). Cox (2001) references studies of auditory imagery by Armstrong *et al.*, (1995) to show that auditory imagery is critical to musical meaning-making, where listeners internally mime the physical actions that they imagine are required to produce specific sounds, thereby accessing new embodied meanings implied within the music. Godøy (1997) offers an account of how auditory imagery exploits the embodied nature of cognition through the metaphorical mapping of embodied schemata and physical gestures during the meaning-making process. Auditory imagery has been described as a key mechanism of meaning-making in sonification, though it is often discussed in terms of information processing (Nees and Walker, 2008; Nees, 2009; Nees and Walker, 2011; Nees and Best, 2013). An account which also explores the potential of auditory imagery for embodied meaning-making may be of use in a sonification design context.

There are definite points of contact between the scientifically oriented sonification research just discussed, and the embodied paradigm illustrated in this thesis. The overall trend shows an acceptance of the embodied nature of certain structures of meaning such as the commonly shared invariant nature of the perceptual responses of psychoacoustics, the perceptual gestalts of ASA and the existence of auditory imagery. However, it could be argued that it also shows a computationalist approach to the process of meaning-making where the listener acts as a processor of embodied symbols. An embodied model, which can be used to create empirically testable sonifications, might open new avenues of meaning-making for scientifically oriented sonification.

1.12 Meaning-making and the Art of Sonification

Vickers (2011; 2012) and Kramer (1993) argue that Peirce's semiotics, especially his theory of indexicality, (see Peirce, 1897/1955) provides a useful conceptual context in which to examine meaning-making in sonification. Brier (2008) argues for a biosemiotics that extends Peirce's original programme to account for embodied meaning-making. Meyer (1956) proposed a similar model of meaning-making in music listening. He relates Peirce's semiotics, where meaning emerges in the interaction between sign, signified and observer, and to Dewey's to concept of lived experience and his conflict theory of emotions, as well as the work of the Gestalt theorists. In this view, daily lived-experience gives rise to stylistic gestalts which the mind draws on to make musical experiences meaningful through the music's fulfilment and denial of expectations related to those gestalts. These ideas bare similarity to those touted by proponents of the embodied mind paradigm, especially enactivism and meaning-making through the metaphorical projection of embodied schemata, as discussed earlier. This suggests that Peirce's semiotics can be approached in sonification research and practice from an embodied point of view when the meanings of the signs involved are recognised as being grounded in, and mediated by, the subject's embodied experience.

Worrall (2013) discusses the historical trend in Western art music towards disembodiment through the gradual reduction of the role of the instrumental performer, to innovations in musical notation, the use of fixed recorded or synthesised sounds as musical material, and a focus on perceptual dimensions of sound over their means of production. He further argues that PMSon inherits this disembodiment through its reliance on music software, where the data acts as the score and the synthesis engine as the instrument without the intermediary of an embodied performer to monitor and adapt the sonic output for the embodied ear in real-time. This results in sonifications which encounter the dimensional entanglements described by the mapping problem (which point is discussed in greater detail elsewhere in this thesis) and being less meaningful to a listener as a result. In the context of sonification, Worrall (2010; 2011) discusses the joint action-perception circuitry of mirror neurons which assigning meaning to the sound by associating perceived sounds to imagined physical sound-making gestures. This represents an embodied approach to meaning-making that parallels the work by Gallese and Lakoff (2005), Cox (2001) and Johnson's original concept of the embodied schemata (Johnson, 1987). A number of researchers have adapted the ideas of Pierre Schaeffer to sonification (Vickers 2012; Diniz *et al.*, 2010; 2012; Parseihian and Katz 2012). Godøy (2006) proposes the gestural sonorous object as an extension of Schaeffer's sound object (*objet sonore*). He argues that Schaeffer's original framework of typological and morphological categories for sound objects are all built around sound producing physical gestures. Godøy builds upon sound object and Smalley's Spectromorphology (Smalley, 1986) to account for embodied meaning-making. Spectromorphology is a descriptive framework for electroacoustic music consisting of detailed categorisation schemes deriving from basic gestural shapes called primal gestures that are extended to add a meaningful low-level organisational structure to musical domains (Smalley, 1997). Graham and Bridges (2015; 2014) argue that the concept of the primal gesture and surrogacy bear resemblance to the concepts of embodied schemata and conceptual metaphor, and that spectromorphology offers an embodied explanation of electroacoustic music. The gestural sonorous object paradigm was considered by Worrall (2013) in the context of using micro-gestural inflections to refine PMSon mapping strategies for the embodied listener. It was also considered by Grond and Hermann (2014) in their development of a set of guidelines for interactive sonification. This represented an introduction of a more embodied approach to meaning-making in Hermann's work.

The embodied nature of meaning-making in music listening has been repeatedly asserted in the field of cognitive musicology with Larson (2010), Brower (2000), Zbikowski (2005), Cox (1999; 2001; 2011) and Johnson (2008) all arguing that the embodied schemata, conceptual metaphors and blends discussed previously are complicit in musical meaning-making, and Kendall (2014) extending these structures to account for embodied meaning-making in electroacoustic music, as discussed earlier.

Diniz *et al.* (2010; 2012) present a framework for interactive sonification which builds on Leman’s work on mediation technologies for embodied music cognition (see Leman, 2008). This framework exploits the embodied nature of musical meaning-making to communicate data to a listener.

There are definite points of contact with the computationalist model of meaning-making in this account of aesthetically oriented sonification, but *en masse* there is a recognition of the need for an embodied model of meaning-making and, in some cases, a direct application of it.

1.13 Discussion

This chapter opened with the argument that an embodied model of meaning-making can provide a unifying theoretical paradigm for sonification which exploits the aesthetic dimensions of sound in an empirically responsible manner.

The analysis of meaning-making in sonification showed that scientific approaches generally tend more strongly towards adopt a computationalist model which treats meaning-making as information processing, than aesthetic approaches. Some of these approaches do still account for embodied dimensions of sound in the information that is processed. This chapter also suggests that embodied approaches to meaning-making tend to be more prevalent in aesthetically oriented sonification research.

An approach to meaning-making in sonification that relies exclusively upon the computationalist model of meaning-making cannot adequately account for the embodied and aesthetic dimensions of meaning-making to which sound is arguably best evolved to support. Computational models limit the scope of meaning making to the processing of sonic symbols but this represents only one possible layer on which the meaning making process can unfold for a listener. In the context of sonification, a designer needs to be able to fully exploit all of a listener’s meaning-making resources in order to design effective sonification solutions that can meet the demands of the specialised tasks for which they are developed. Taking embodied modes of meaning-making into account might open up new dimensions for effective communication with a listener offering the sonification designer a wider range of tools for communicating with a listener via sound.

An embodied approach might also help to enrich and further support the computational levels of meaning making that can be leveraged in sonification by providing a familiar grounding context for the interpretation of sonic symbols. When symbols in a sonification are not well grounded, related to the data they symbolise in an obvious way, confusion is created for the listener. The problem can be addressed through training where the listener learns what the sonic symbols, and the translations they undergo, are intended to mean. This solution requires the listener to consciously process each successive sonic symbol and relation during sonification listening, until their meanings have been internalised. Such an approach puts three levels of cognitive load on the listener where first they must learn the meanings before ever using a sonification, secondly remember the meanings of each symbol while listening, and lastly they must correctly apply those meanings in a real-time scenario. None of these tasks are trivial as human subjects, being more suited to intuitive modes of meaning-making, are thought to be slow and cumbersome when processing symbols and information by rule (See Kahneman, 2011). This is especially true in the current age of ubiquitous and invisible smart computing (Bibri, 2015). This might not pose a problem for specialist sonification solutions, which tend to require listener training. It does pose a problem for general use by large non-specific audiences of online listeners, the sonification listenership considered in this thesis. Overlooking embodied meaning-making runs the risk of reducing the accessibility of sonification to these wide general audiences who are not in a position to engage in perquisite training activates before listening to a sonification. This is a problem that must be addressed if sonification is to become a mass medium as Barrass (2012) advocates, or to become a standard technique for creating value for a listener from the hidden dataflows of the internet of things (Rimland *et al.*, 2013).

Another limit of the computationalist paradigm highlighted by Johnson (Johnson, 2010) was foreshadowed earlier in this chapter. The paradigm does not allow for any of the modes of meaning-making that are classically categorised under the heading of aesthetics. He takes this further by appeal to Dewey’s argument that ordinary experience is composed of multitudes of aesthetic dimensions, which are critical to meaning-making. As a result, the embodied component of human meaning has aesthetic dimensionality, a position that is revisited in Chapter 3 of this thesis. Computationalism then cannot account for the vast majority of aesthetically fused human meaning. This makes it a particularly limited paradigm for sonification, an area of research and practice that is chiefly concerned with meaning and meaning-making. It also presents a barrier to the widespread adoption of sonification as a mass medium as Barrass (2012) argues that a focus on aesthetics is crucial to turning sonification into a mass medium.

PMSon and the use of environmental and musical sound are isolated in this chapter as sonification techniques that could support an embodied approach to meaning-making. These were chosen because they may hold the potential to exploit the listeners' embodied meaning-making faculties. In order to achieve this a more in depth understanding of how to create effective mapping strategies for embodied sonification approaches is required. These mapping strategies are explored in detail in Chapters 3, 4 and 5. Before these mapping strategies can be explored however, theoretical tools for conceptualising and working with the embodied dimensions of sound and listening practices in a sonification context are required. Chapter 2 of this thesis takes up this task. It provides a sonification listening model and some novel conceptual measurement schemes for working with sound in a sonification context. The rest of the thesis applies the model and conceptual measures to explore a number of embodied frameworks and strategies for creating communicatively effective sonifications which exploit the aesthetic dimensions of sound made available by embodied meaning-making.

Chapter 2: The Embodied Sonification Listening Model

2.1 Introduction

This chapter opens with a short discussion of sonification listening. The Embodied Sonification Listening Model (ELSM) is then introduced. This model describes the sonification listening experience in terms of the embodied cognitive faculties involved. The presentation of the ESLM is followed by four hypotheses describing how listeners draw on their embodied schematic knowledge to interpret changes in pitch and tempo in sonification listening. An empirical listener evaluation intended to test these hypotheses, and a discussion of the findings obtained and their relevance to the ESLM are then presented.

2.2 The Embodied Sonification Listening Model

Chapter 1 concluded that a model of the embodied meaning-making faculties active in sonification listening is required to guide the design of communicatively effective sonification mapping strategies. Vickers (2012) makes the similar argument that the modes of listening proposed by thinkers like Schaeffer (1966), Chion (1994), and Gaver (1989) are insufficient in describing sonification listening and calls for a new paradigm that is exclusively focused on describing the richness and diversity of the sonification listening experience.

This chapter introduces the Embodied Sonification Listening Model, Figure 2.1, which describes how meaning emerges in sonification listening, from an embodied perspective. Typically, a listener does not have direct access to the data or the original data source being represented during sonification listening. As a result they must construct an imaginary model of the data on the basis of the cues and directions provided by the sonification. In the same way that a sonification designer creates a mapping strategy from data to sound, the listener must create a cognitive mapping strategy from that sound back to an imagined data source. The embodied sonification listening model provides a theoretical explanation of the embodied meaning-making faculties involved in this process. It relies on the embodied meaning-making faculties discussed in detail in Chapter 1 to describe the sonification listening process. The ESLM involves two novel conceptual measurement schemes, the embodied sonic dimension and embodied sonic complex. These were devised to account for traditional dimensions of sound such as pitch, duration, amplitude and timbre, and also to account for the aesthetic dimensions of sound, examined in this thesis, which are too complex to be adequately described by these standard dimensions alone. Examples of such dimensions might be a sense of tension as communicated in prosodic information of human vocalisations or the unique sense of place established by a specific soundscape. These conceptual measurement schemes are also intended to address the perceived need for dedicated theoretical descriptors for sonification (Walker and Nees, 2011). They were motivated by Koestler's concepts of the holon and the holarchy (see Koestler, 1967) and must be defined and understood before the model can be described effectively. A holon is something which is simultaneously a whole and a part of a larger whole while a holarchy is a hierarchical arrangement of individual holons. An embodied sonic dimension is defined here as any individual sonic aspect that a listener can attend to as a meaningful perceptual unit and which retains a single conceptual identity, while evolving in time along a continuous bi-polar axis. An embodied complex is defined as any perceptual grouping that contains multiple embodied sonic dimensions and can also be attended to by a listener as a meaningful perceptual unit, with a single conceptual identity.

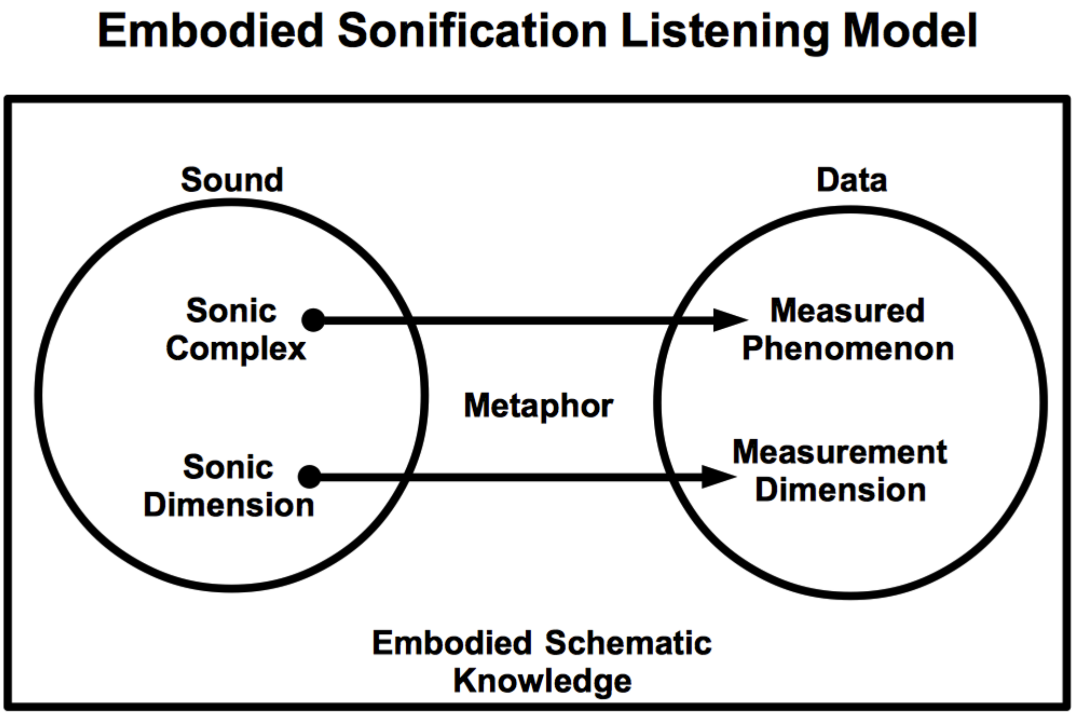


Figure 2.1. The Embodied Sonification Listening Model

The model asserts that listeners attend to the sound as though it were the data during sonification listening. Thus, the sound is experienced as a metaphor for the data. This process unfolds against a background context of embodied schematic knowledge which grounds the meaning of the sounds. Embodied schematic knowledge refers to a listener's everyday knowledge of their physical, social and cultural environments, as represented in, and mediated by the embodied schemata discussed in Chapter 1. This embodied schematic knowledge can be leveraged as a means of grounding perceptual and conceptual symbolic content via the processes of cross domain mapping, conceptual metaphor and conceptual blending, which are also described in greater depth in Chapter 1. In the case of sonification listening the listener's background of embodied schematic knowledge contains their understanding of the sound, the data, any instructions or training they have received regarding the sonification and any associations, conscious or unconscious, the listener draws between or to these elements. This knowledge determines the cognitive mapping strategy a listener employs to map the sound back to an imagined data source during sonification listening. There are two metaphorical mappings depicted in the ESLM.

Firstly, the listener maps the embodied sonic complex of the sonification to the source of the data i.e., the phenomenon of which the data is a measure. Secondly, the listener also maps changes along embodied sonic dimensions within that complex to changes along the measured dimensions of the data i.e., the dimension along which the phenomenon is measured. For example, where litres of water are represented using a pitch-mapped sine tone, the sine tone (embodied sonic complex) is a metaphor for the water and the changes in pitch (embodied sonic dimension) are a metaphor for changes in volume. Where wind-speed data is mapped to the perceived RPM of an imaginary engine, the engine sound (embodied sonic complex) is a metaphor for the wind and changes in RPM (embodied sonic dimension) are a metaphor for changes in the speed of that wind.

As noted previously, the listener's background of embodied schematic knowledge contains their understanding of the sound, the data and any instructions or training they have received for the sonification. This knowledge is grounded in the listeners embodied experience through embodied schemata and these embodied schemata determine how the embodied sonic dimensions are mapped to data. A similar phenomenon is referred to by Walker (2002) as polarity and is discussed in more detail shortly. For example when the speed of a train is mapped to the sound of flowing water, an increase in the speed of the water flow (embodied sonic dimension) is likely to coincide with an increase in the speed of the train as both share a common measure, speed, which is structured by the Fast-Slow schema (see Johnson, 1987). When the depth of a submarine is mapped to pitch, a decrease in pitch (embodied sonic dimension) is likely to correspond to an increase in depth. This is because both depth and pitch are structured by a common Up-Down schema (see Johnson, 1987; Zbikowski, 2005). For depth however, an increase in the data means downward motion and so a decrease in pitch might be interpreted as an increase in data. In this case, the listener's embodied schematic knowledge of the data determines their experience of the sonification.

The ESLM is proposed as a tool for guiding the design of sonifications that can exploit the embodied nature of meaning-making during sonification listening and it is employed as such throughout the remainder of this thesis. This chapter examines and provides empirical support for the role of embodied schematic knowledge in sonification listening, as described in the ESLM. Chapter 3 explores and provides empirical support for the sonic dimensions postulated in the model. Chapter 4 explores and provides empirical support for the sonic complex proposed by this model. Chapter 5 closely examines and provides empirical support for the role of conceptual metaphorical mappings in sonification listening. Chapter 6 summarises all of the empirical support that this thesis provides for the ESLM.

2.3 Embodied Schematic Knowledge in Sonification Listening

As previously noted the ESLM suggests that sonification listening may be reliant on the embodied meaning-making faculty of conceptual metaphorical mapping and that the cognitive strategies involved in associating the sound with the data depend on the listener's previous embodied schematic knowledge. The remainder of this chapter explores how this knowledge influences the cognitive mapping strategies listeners use to interpret pitch and tempo dimensions during sonification. It presents a number of hypotheses about the relationship between embodied schematic knowledge and sonification listening, which are then evaluated empirically.

The embodied cognition literature on conceptual metaphorical mapping states that listeners draw on previous domains of everyday embodied experience to make sense of new experiences (Lakoff and Johnson, 1980). The ESLM builds upon this and suggests that a listener's previous embodied schematic knowledge influences how they will interpret a sonification during sonification listening. The model suggests that when listening to abstract sounds, sounds for which no extraneous referential context is provided (see Chapter 1, section 1.8) listeners will identify embodied schematic patterns in the sound and interpret the sound in relation to those embodied schemata. It is also suggested that a listener's knowledge of the data source for a sonification will help to determine which embodied schematic knowledge they will draw upon to interpret the sonification.

Walker (2000) discusses the concept of polarity in regards to the cognitive mapping strategies that listeners employ to interpret a sonification. A positive-polarity mapping occurs when increases in a sonic dimension such as pitch or tempo are interpreted as increases in the value of the data and decreases in the dimension are interpreted as decreases in the data. A negative-polarity mapping occurs when a decrease in a sonic dimension is interpreted as an increase in the value of the data and an increase in the dimension is interpreted as a decrease in the data. It is here suggested that the domain of embodied schematic knowledge a listener draws from to interpret a sonification will help to define the polarity of the cognitive mapping strategy the listener employs.

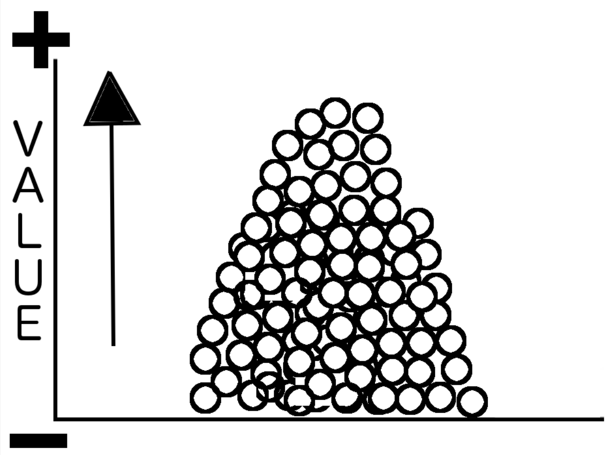


Figure 2.2. Embodied Experience of Multiple Entities in a Collection

In everyday embodied experience, increases and decreases in the number of objects in a collection tend to coincide directly with increases and decreases in vertical height (Lakoff and Johnson, 1980; 1999; Lakoff and Núñez, 2000) (see Figure 2.2). For example, as rice pours into a saucepan, the grains pile on top of one another and the height of the rice increases as the number of grains in the container increases. Johnson (1987) argues that experiences such as this give rise to an Up-Down schema. Zbikowski (1997) and Brower (2000) have suggested that listeners draw upon their embodied schematic knowledge of vertical height, in the form of this Up-Down schema, to interpret pitches in music. They argue that this provides the basis for the conceptualisation of pitches in terms of high and low. This would suggest that when knowingly attending to a sonification of number data, listeners would relate their embodied schematic knowledge of spatial entities to their knowledge of pitch as vertical height and apply a positive-polarity mapping to interpret increasing pitch as increasing value and decreasing pitch as decreasing value. The listener must know that a sonification communicates number data in order for them to interpret it in this manner because it is their embodied schematic knowledge of the data that determines their cognitive mapping strategy and defines how they interpret a sonification.

Increasing the tempo at which sonic stimuli are presented within a given timeframe, translates to increasing the number of sonic stimuli presented within that timeframe. Listeners are likely to apply a positive-polarity mapping and interpret increases in the number of stimuli presented in a sonification with increases in the number of phenomena represented in a data set. Johnson (2005) and Lakoff and Núñez (2000) argue that the collection schema provides the embodied schematic basis for this understanding of numbers where a collection consists of several entities grouped together and considered as a whole. On these bases, the following hypotheses are proposed:

(A1) When knowingly listening to a sonification of number data, listeners draw upon their embodied schematic knowledge of entities in collections to interpret changes in pitch using a positive-polarity mapping.

(A2) When knowingly listening to a sonification of number data, listeners draw upon their embodied schematic knowledge of entities in collections to interpret changes in tempo using a positive-polarity mapping.

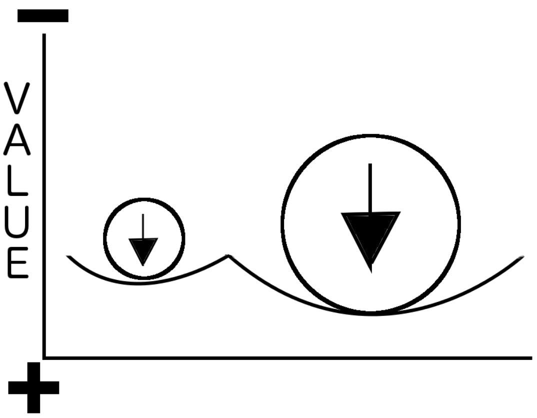


Figure 2.3. Embodied Experience of Entities in Spatial Fields

In embodied experience increases in an entity's physical attributes, e.g., weight, volume and size often correlate with increases in the mass of the entity. This in turn correlates with increased downward force within the gravitational field and a decrease in vertical position in the spatial field (Lakoff and Johnson, 1999). It is has been argued that subjects interpret a large number of attributes in this manner due to a conceptual metaphorical mapping which frames individual attributes as individual entities (Lakoff, 1994). When sonified data is thought by the listener to represent an attribute value, the listener will likely draw from their embodied schematic knowledge of entities in spatial fields to understand the sonification. This means that listeners are likely to apply a negative-polarity mapping and interpret increases in pitch as decreases in value and vice-versa when listening to sonifications of attribute data.

Lakoff and Johnson (Lakoff and Johnson, 1999) argue that subjects conceptualise time in terms of physical space. Johnson (Johnson, 2008) and Zbikowski (Zbikowski, 2005) extend this to music, arguing that listeners interpret musical motion, and tempo, in terms of space. This suggests that listeners might also interpret tempo in terms of space during sonification listening. As the tempo of a repeated sonic pattern decreases the amount of temporal space between each element of the pattern increases. As such, a decrease in tempo relates to an increase in space, and an increase in tempo correlates to a decrease in space. This would suggest that when knowingly attending to a sonification of attribute data, listeners would relate their embodied schematic knowledge of spatial entities to tempo and apply a negative-polarity mapping to interpret decreasing tempo as increasing value and increasing tempo as decreasing value. On these bases, the following hypotheses are proposed:

(B1) When knowingly listening to a sonification of attribute data, listeners draw upon their embodied schematic knowledge of entities in spatial fields to interpret changes in pitch using a negative-polarity mapping.

(B2) When knowingly listening to a sonification of attribute data, listeners draw upon their embodied schematic knowledge of entities in spatial fields to interpret changes in tempo using a negative-polarity mapping.

Listeners need to know what data is being sonified when they listen to a sonification in order to access and apply the domain of embodied schematic knowledge which associates the sound with the original data-source in the cognitive mapping strategy they use to make the sonification meaningful. If a designer sonified attribute data, but a listener believes they are listening to number data, they will interpret the sonification in reference to their embodied schematic knowledge of amounts. As such, the designer needs to find a way to inform the listener of the kind of data they are hearing whether by informing them before-hand or preferably, communicating the data type in the sonification.

An interesting example of real-world counterparts to the attribute mapping strategy for pitch is provided by the example of water pouring into a bottle. As water pours into a bottle an audible rise in pitch is heard as the air cavity within the bottle is reduced in size. However this process also produces a second much softer sound of decreasing pitch as the mass of the water increases. The sound produced by the increasing water mass is usually masked by the sound of the decreasing air cavity. A listener who keenly observed the bottle while attending carefully to the sound would perceive the rising tone associated with the decreasing air cavity and the falling tone the tone associated with the increasing water mass. Both the sounds created by the decreasing cavity and increasing water mass are consistent with negative-polarity pitch mappings described in hypothesis B1.

Shepard tones represent an interesting auditory expression of the Up-Down schema in which two sine waves are arranged in octave relationships. A Shepard scale is composed of a configuration of Shepard tones separated by semitone steps. The amplitude of each tone in the scale is a Gaussian function of its proximity to a base frequency. By continually increasing this base frequency creates the illusion of a scale of perpetually increasing pitch while decreasing it creates the illusion of perpetually decreasing pitch. The Shannon Portal project mapped the movement of people through Shannon airport to the panning orientation, amplitude and pitch direction of a Shepard scale in the context of an auditory display installation (Fernström and Brazil, 2009). A Shepard scale might be used to give the user a sense of perpetually increasing or perpetually decreasing data values making the Shepard scale useful for bypassing the limits of human frequency perception in a sonification context.

2.4 Experimental Investigation

Four hypotheses about the relationship between embodied schematic knowledge and the listeners' cognitive strategies for interpreting changes in pitch and tempo during sonification listening are laid out above. An empirical listener evaluation was undertaken to test the validity of these four hypotheses. Participants were recruited online through the crowdsourcing platform Crowd Flower[6]. Online evaluation methods have been previously used in an auditory display context by Walker (Walker, 2000). The limitations of crowdsourcing are addressed in the following section, section 2.4.1. Each of the evaluations was designed, hosted and delivered on the Survey Gizmo[7] web-platform. Survey Gizmo is a reliable web-platform for creating and delivering evaluations and surveys over the internet. All participants were required to pass a rigorous validation test to prove that they were undertaking the evaluations using a stereo setup with either a good set of headphones or a double speaker array. Potential participants who did not pass the validation test were not allowed to take part in the experiment. The study opened internationally to ensure that the results were not specific to a particular culture. There were a total of 112 participants from 34 countries This number consisted of 26% females and 74% males. 26% of listeners had formal musical training and 22% played an instrument. A more detailed demographic breakdown is presented in appendix 2.2.

Twelve individual stimuli of 30 seconds length each were used in the experiment. Six stimuli exhibit linear variation in the pitch of a sine tone over time and the other six exhibit linear variations in the tempo of concurrently presented sine bursts of 440Hz frequency over time. Three of the pitch stimuli, the increasing pitch stimuli, begin at 440Hz and rise to 1760Hz over the course of playback. The other three, the decreasing pitch stimuli, begin at 1760Hz and fall to 440Hz. Three of the tempo stimuli, the increasing tempo stimuli, begin at 120bpm and rise to 600bpm. The other three, the decreasing tempo stimuli, begin at 600bpm and fall to 120bpm. All of the sounds were designed and created using the Csound audio programming language. Appendix 2.1

---

[6] www.crowdflower.com

[7] www.surveygizmo.com

contains example evaluation questions. The stimuli used, the entire set of questions and the raw data collected during this evaluation are included in the digital appendices for this chapter on the appendices DVD.

The data types used in the experiment are listed in Table 2.1. It is expected that increases in pitch and tempo for thickness, volume, strength, stiffness, hardness, size, mass, surface area, empty space, depth and angle will all be interpreted using cognitive mapping strategies that have a negative-polarity by the listener. It is expected that increases in pitch and tempo for number of students, number of ants, concentration, population, absorbency, stock count, tension, acceleration, temperature, RPM and speed will all be interpreted using cognitive mapping strategies that have a positive-polarity by the listener. The data types and their expected polarities are presented in Table 2.1. Production rate and crop yield were also explored in this experiment in order to determine what kind of cognitive mapping strategy a listener might use to interpret them.

Negative-Polarity	**Positive-Polarity**	**Uncertain**
Thickness	Number of Students	Production Rate
Volume	Number of Ants	Yield
Strength	Concentration	
Stiffness	Population	
Hardness	Absorbency	
Size	Stock Count	
Mass	Tension	
Surface-Area	Acceleration	
Empty-Space	Temperature	
Depth	RPM	
Angle	Speed	

Table 2.1. *Expected Interpretation Strategies per Data Type*

2.4.1 Limitations of Crowdsourcing

Oh and Wang (2012) discuss crowdsourcing and web based evaluations in the context of music perception experiments. They cite recent research into the limitations of crowdsourcing for empirical testing. They discuss how the quality of crowdsourced data can be affected by a number of factors. Malicious users can purposely submit compromised data which will affect the outcome of an evaluation. Lazy users can put in the minimal amount of effort required to fulfil the basic requirements of a specific task. Some users may become distracted by factors in their own environments and listeners using sub-standard hardware may not be able to hear the stimuli properly. Measures were put in place in the evaluations discussed in this thesis to account for these users. Listeners were chosen from Crowdflower's internal pool of trusted level three contributors. Crowdflower evaluates the strength of its own contributors through rigorous test evaluations and those who have achieved a high level of accuracy in testing are awarded level three status[8]. These contributors provide the highest quality answers for crowdsourced evaluations. All of the evaluations undertaken in this thesis used level three contributors. This choice was made to limit the number of malicious, lazy, distracted and low quality hardware listeners. Time limits were imposed on each evaluation to make sure that any participants who sped through the evaluation too quickly, or who crawled through the evaluation too slowly were eliminated. Listeners who were idle for long periods of time, relative to the kinds of tasks presented in the respective evaluations, were also eliminated. A system was implemented to ensure that listeners had access to stereo playback. Audio was played through both stereo channels and listeners were asked to identify the sequences of numbers contained in the audio. Any listener who could not identify the numbers correctly was prevented from taking part in the evaluation. Listeners were also asked if the listening equipment they were using was of good quality and if not they were excluded from the evaluations. Furthermore, listeners were asked to confirm that they were undertaking the evaluations using good quality stereo speakers or a good quality set of headphones. Listeners who did not confirm were eliminated from the evaluation. A mobile friendly version of each evaluation was provided for those listeners accessing the evaluations by tablet or phone. Listeners were allowed to access the evaluations using mobile technologies because it is hoped that the results generated might be of use to the design of effective sonifications in the context of mobile smart technologies and the internet of things as discussed in the introduction of Chapter 1.

Another problem facing the evaluations in this thesis was the participation of musically trained listeners. Research has shown that listeners with musical training can interpret sonifications differently to listeners without musical training (Neuhoff *et al.*, 2002; Sándor and Lane, 2003). The research in this thesis is intended help establish sonification as a tool for use by large non-specific audiences of international listeners regardless of their musical background. Audiences of this nature will necessarily contain listeners with differing degrees musical training. Each evaluation sourced listeners from a large international pool of non-specific participant and as such the number of musically

---

[8] http://crowdflowercommunity.tumblr.com/post/108559336035/new-performance-level-badge-requirements

trained listener would be expected to represent audiences of this type. During each evaluation listeners were asked if they played an instrument or had formal musical training and the levels to which they were engaged in either activity. The numbers of listeners with musical expertise fell roughly within the range of 20% to 23% while the number of listeners with instrumental skills well within the range of 16% to 36%. These figures were deemed acceptable as studies in the field tend to show a large variance in terms of the numbers of musically trained listeners considered. For example in their exploration of musicians understanding of pitch change in a sonification *vs.* non-musicians only 28% of Neuhoff *et al.*'s (2002) listeners were musically trained while a in similar study by van Zuijen *et al.* (2005) 48% of listeners had musical training. The exact numbers of musically trained listeners are provided in the demographics breakdowns for each evaluation in the appendices for Chapters 2, 3, 4 and 5.

2.5 Experimental Procedure

Listeners take part in a short training exercise to familiarise themselves with the procedure before beginning. They were instructed to listen to the sounds as many times as needed before answering a question. Listeners were presented with each of the stimuli described in the previous section twice. On the first presentation, listeners were told that a stimulus represents a specific data type and asked to categorise the data value represented as increasing or decreasing. On the second presentation they were told that the stimulus represents a different data type and were asked again to categorise the data value represented as increasing or decreasing. These stimuli were not presented consecutively. The impact of changing the listener's knowledge of the data type in a sonification can be measured on this basis.

2.6 Results

Pitch Stimulus	Stimulus Direction	Data-Type	Increasing	Decreasing
A	Decreasing	Number of Students	24%	76%
		Surface Area	31%	69%
B	Decreasing	Thickness	48%	52%
		Acceleration	19%	81%
C	Increasing	Volume	68%	32%
		Temperature	83%	17%
D	Increasing	Empty Space	61%	39%
		Amount of Ants	82%	18%
E	Increasing	Concentration	83%	17%
		RPM	87.5%	12.5%
F	Decreasing	Mass	35%	65%
		Population	20.5%	79.5%

Table 2.2. *Results for Pitch Stimuli*

Tempo Stimulus	Stimulus Direction	Data-Type	Increasing	Decreasing
A	Decreasing	Strength	35%	65%
		Stiffness	38%	62%
B	Decreasing	Depth	51%	49%
		Absorbency	17%	83%
C	Increasing	Production-Rate	60%	40%
		Angle	63%	37%
D	Increasing	Hardness	64%	36%
		Stock Count	82%	18%
E	Increasing	Yield	59%	41%
		Tension	81%	19%
F	Decreasing	Size	29.5%	70.5%
		Speed	29%	71%

Table 2.3. *Results for Tempo Stimuli*

The results for pitched stimuli are presented in Table 2.2 and the results for tempo stimuli are presented in Table 2.3. The results were further analysed using McNemar tests on the twelve pairs of stimuli, comparing categorisations of increasing and decreasing value for each data type (see McNemar, 1955; Eliasziw and Donner, 1991). A McNemar test compares the differences between two sets of numbers to determine if they are significant. Because listeners were provided with a single stimulus and only the data type is changed the McNemar test help to determine if the listener's knowledge of the data type can significantly effect their interpretation of a stimulus. These inferential statistics are presented below alongside a consideration of the descriptive statistics for each question.

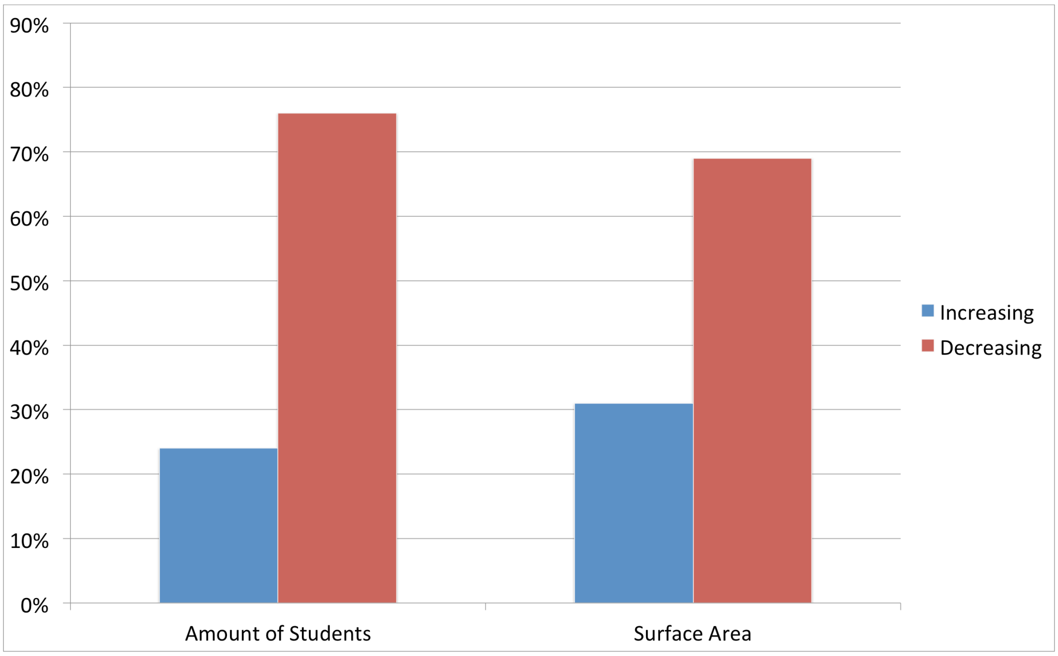


Figure 2.4. Results for Pitch Stimulus A

Pitch stimulus A decreased in pitch and the majority of listeners categorised the stimulus as decreasing in value for both data types as illustrated in Figure 2.4 and recorded in Table 2.2. 76% of listeners categorised the stimulus as representing decreasing value for number of students. Only 31% categorised an increase in value for surface area. This is a relatively small number in comparison to the other results obtained in this experiment. The McNemar test results for pitch stimulus A showed that listeners interpretation of the stimulus did not differ significantly when the description of the data-type was changed from number of students to surface area X^2 (1, N = 112)=1.6, $p>.05$, phi = 0.11, the odds ratio is .666. This suggests that the majority of listeners interpret number of students and surface area using a positive-polarity mapping and that the number of listeners who interpret surface area with a negative-polarity mapping is not of substantial practical difference to the number of listeners who interpret number of students with a negative polarity-mapping.

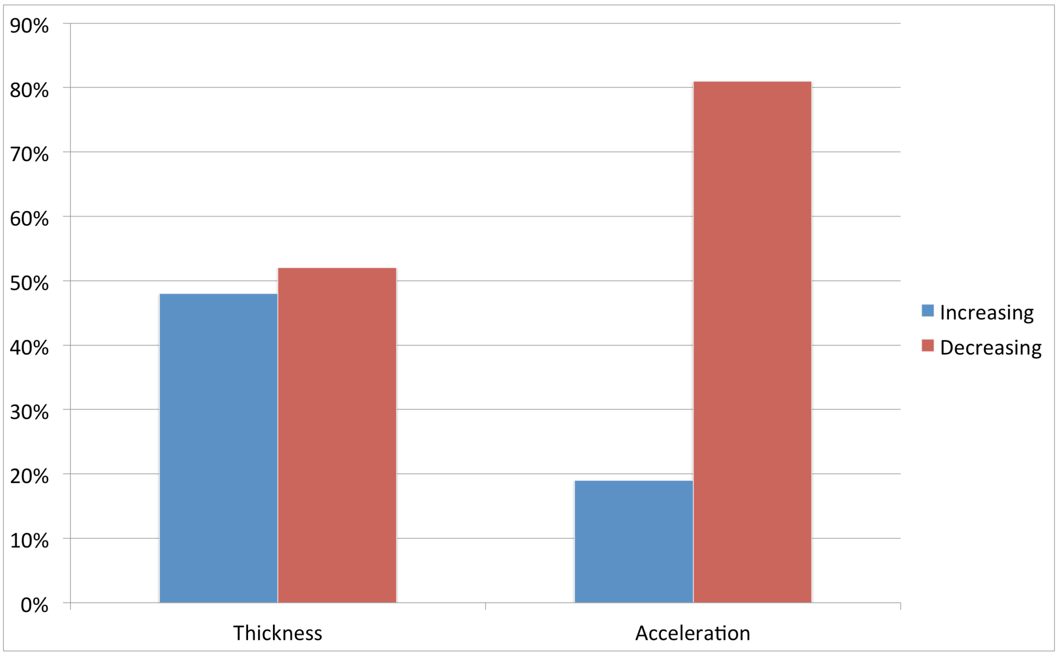


Figure 2.5. Results for Pitch Stimulus B

Pitch stimulus B decreased in pitch and the majority of listeners categorised the stimulus as decreasing in value for both data types as illustrated in Figure 2.5 and recorded in Table 2.2. 52% of listeners categorised the stimulus as decreasing in value for thickness data. This is a relatively small percentage in comparison to the other results obtained in this experiment. The McNemar test results for pitch stimulus B showed a statistically significant number of people interpreted the stimulus differently when the description of the data-type was changed from a measure of thickness to acceleration X^2 (1, N = 112)=19.1, $p<.001$, phi = .41, the odds ratio is 3.75. The effect size represented by phi is medium. This suggests that the majority of listeners interpret thickness and acceleration using a positive-polarity mapping but there is a medium sized tendency amongst listeners to interpret thickness with a negative-polarity mapping in comparison to acceleration.

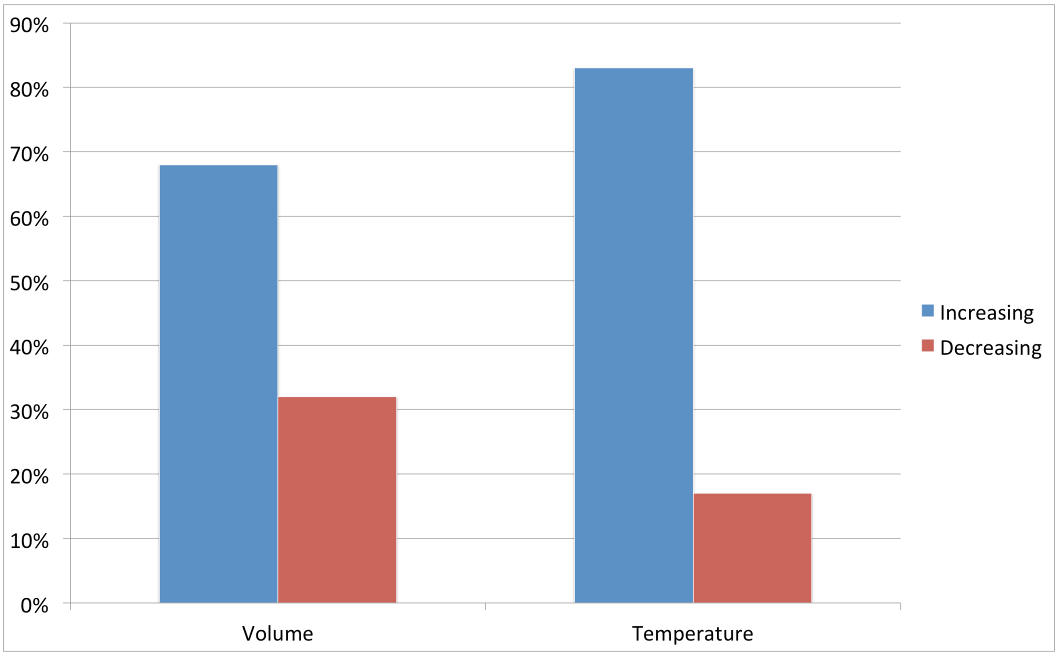


Figure 2.6. Results for Pitch Stimulus C

Pitch stimulus C increased in pitch and the majority of listeners categorised the stimulus as increasing in value for both volume and temperature data as illustrated in Figure 2.6 and recorded in Table 2.2. The results for pitch stimulus C showed a statistically significant number of people interpreted the stimulus differently when the description of the data-type was changed from volume to temperature X^2 (1, N = 112)=7.41, $p<.01$, phi = .25, the odds ratio is 2.54. The effect size represented by phi is small suggesting that this effect is weak. This suggests that the majority of listeners interpret volume and temperature using a positive-polarity mapping but there is a small tendency amongst listeners to interpret volume with a negative-polarity mapping in comparison to temperature.

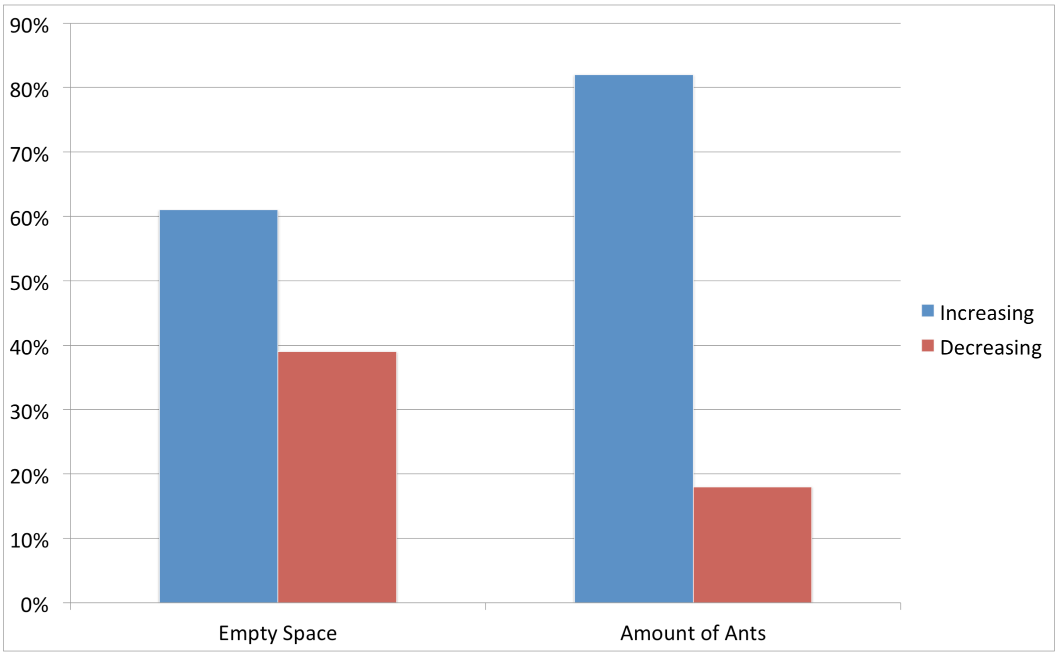


Figure 2.7. Results for Pitch Stimulus D

Pitch stimulus D increased in pitch and the majority of listeners categorised the stimulus as increasing in value for both empty space and number of ants as illustrated in Figure 2.7 and recorded in Table 2.2. 61% of listeners categorised the stimulus as increasing in value for empty space. This is a relatively small percentage in comparison to the other results obtained in this experiment. The McNemar test results for pitch stimulus D showed a statistically significant number of people interpreted the stimulus differently when the description of the data-type was changed from a measure of empty space to a measure of a number of ants X^2 (1, N = 112)=16.94, $p<.001$, phi = .38, the odds ratio is .17. The effect size represented by phi is medium suggesting that this effect is of medium strength. This suggests that the majority of listeners interpret empty space and number of ants using a positive-polarity mapping but that there is a medium sized tendency amongst listeners to interpret empty space with a negative-polarity mapping in comparison to number of ants.

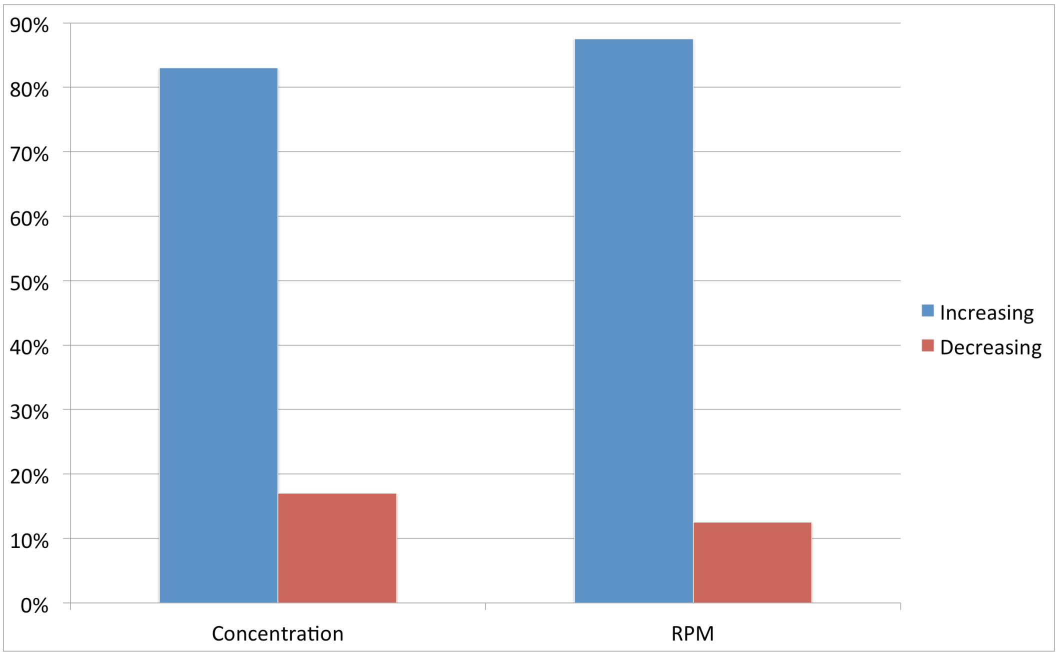


Figure 2.8. Results for Pitch Stimulus E

Pitch stimulus E increased in pitch and the majority of listeners categorised the stimulus as increasing in value for both concentration and RPM data as illustrated in Figure 2.8 and recorded in Table 2.2. Large numbers of listeners categorised the stimulus as increasing for both concentration and RPM. The McNemar test results for pitch stimulus E, showed that the number of people who interpreted the stimulus differently when the data-type was changed from concentration to RPM did not differ to statistically significant degree X^2 (1, N = 112)=1.31, p>.05, phi = .1, the odds ratio is 1.71. This suggests that the majority of listeners interpret concentration and RPM using a positive-polarity mapping and that the number of listeners who interpret concentration with a negative-polarity mapping is not of substantial practical difference to the number of listeners who interpret RPM with a negative polarity-mapping. Relatively large minorities of listeners interpret both concentration and rpm with a positive-polarity mapping in comparison to the other data collected in this evaluation.

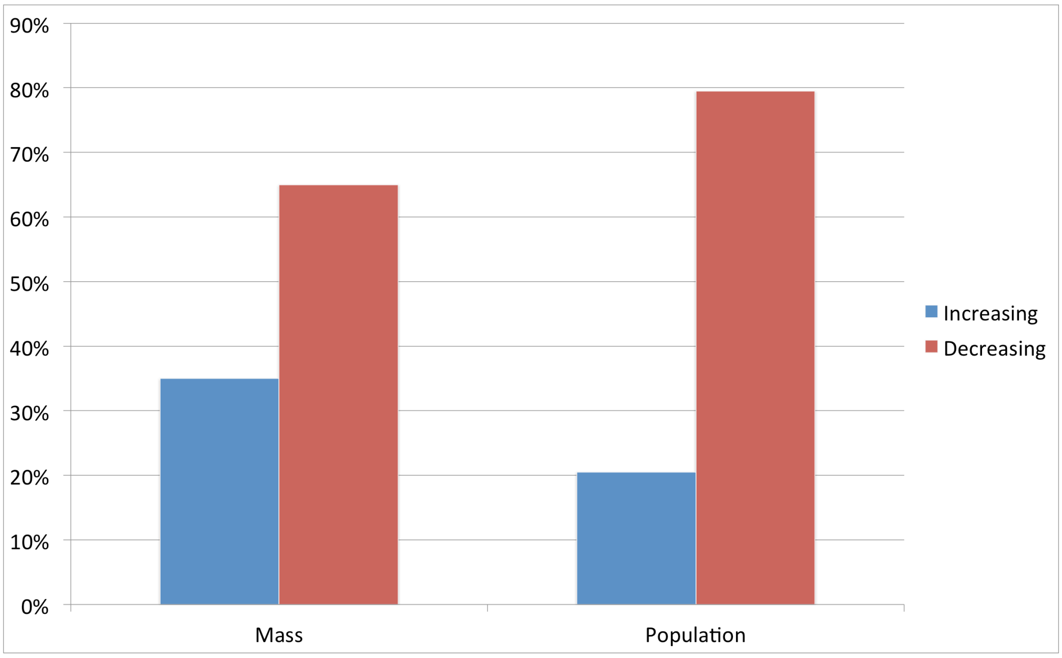


Figure 2.9. Results for Pitch Stimulus F

Pitch stimulus F decreased in pitch and the majority of listeners categorised the stimulus as decreasing in value for both mass and population data as illustrated in Figure 2.9 and recorded in Table 2.2. 65% of listeners categorised the stimulus as decreasing in value for mass. This is a relatively large percentage in comparison to the other results obtained in this experiment. The McNemar test results for pitch stimulus F, showed a statistically significant number of people interpreted the stimulus differently when the description of the data-type was changed from a measure of mass to population X^2 (1, N = 112)=7.11, p<.01, phi = .25, the odds ratio is .38. The effect size represented by phi is small suggesting that this effect is of a weak strength. This suggests that the majority of listeners interpret mass and population with a positive-polarity mapping but that there is a small tendency amongst listeners to interpret mass with a negative-polarity mapping in comparison to population.

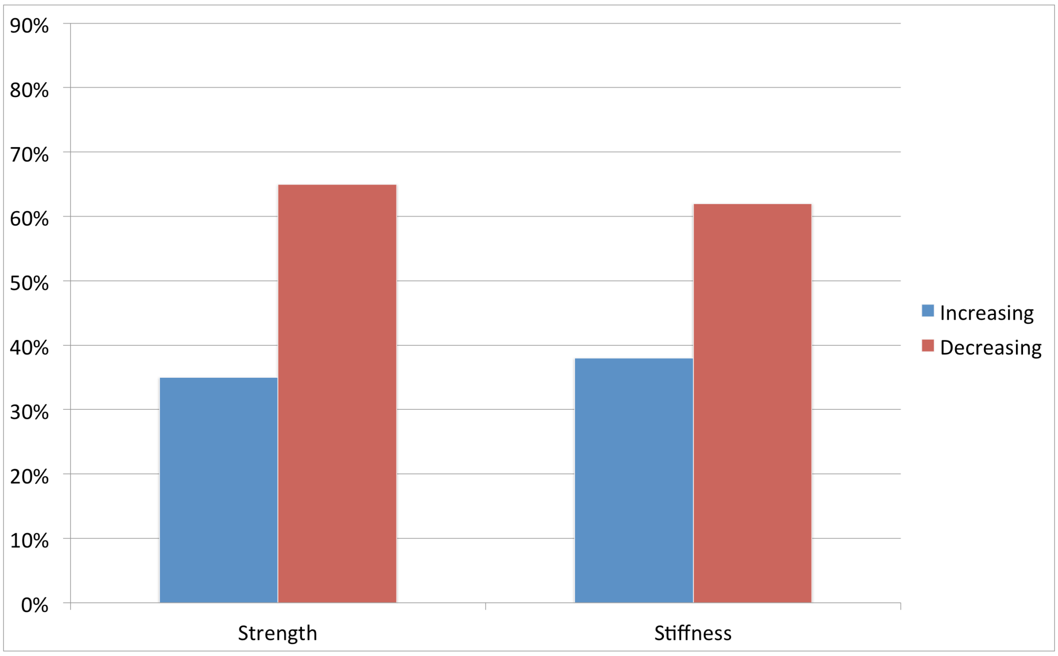


Figure 2.10. Results for Tempo Stimulus A

Tempo stimulus A decreased in tempo and the majority of listeners categorised the stimulus as decreasing in value for strength and stiffness data as illustrated in Figure 2.10 and recorded in Table 2.3. 65% of listeners categorised the stimuli as decreasing for strength and 62% categorised it as decreasing for stiffness. Both of these figures are relatively small compared to the other results obtained. The McNemar test results for tempo stimulus A, showed that the number of people who interpreted the stimulus differently when the data-type was changed from strength to stiffness did not differ to a statistically significant degree X^2 (1, N = 112)=.3, *p*>.25, phi = 0.05, the odds ratio is .82. This suggests that the majority of listeners interpret strength and stiffness using a positive-polarity mapping and that the number of listeners who interpret strength with a negative-polarity mapping is not of substantial practical difference to the number of listeners who interpret stiffness with a negative-polarity mapping. Large minorities of listeners interpret both strength and stiffness with a negative-polarity mapping in comparison to the other data collected in this evaluation.

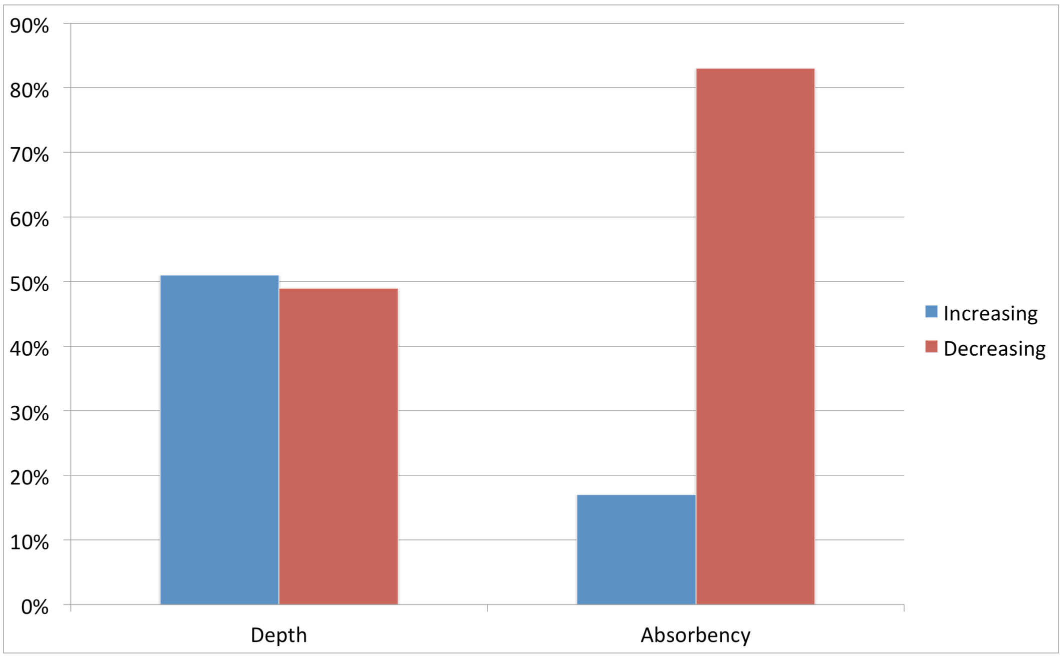


Figure 2.11. Results for Tempo Stimulus B

Tempo stimulus B decreased in tempo and the majority of listeners categorised the stimulus as decreasing for absorbency but 51% of listeners categorised it as increasing for depth data as illustrated in Figure 2.11 and recorded in Table 2.3. This is a relatively large figure in comparison to the other results obtained in this evaluation. The McNemar test results for tempo stimulus B, showed a statistically significant number of people interpreted the stimulus differently when the data-type was changed from depth to absorbency X^2 (1, N = 112)=26.7, $p<.001$, phi = .48, the odds ratio is 5.75. The effect size represented by phi is large suggesting that this effect is strong. This suggests that the majority of listeners interpret absorbency with a positive-polarity mapping but that roughly half of the listeners tended to interpret depth with a negative-polarity mapping. The results also suggest that the number of listeners interpreting depth with a negative-polarity mapping is large in comparison to absorbency.

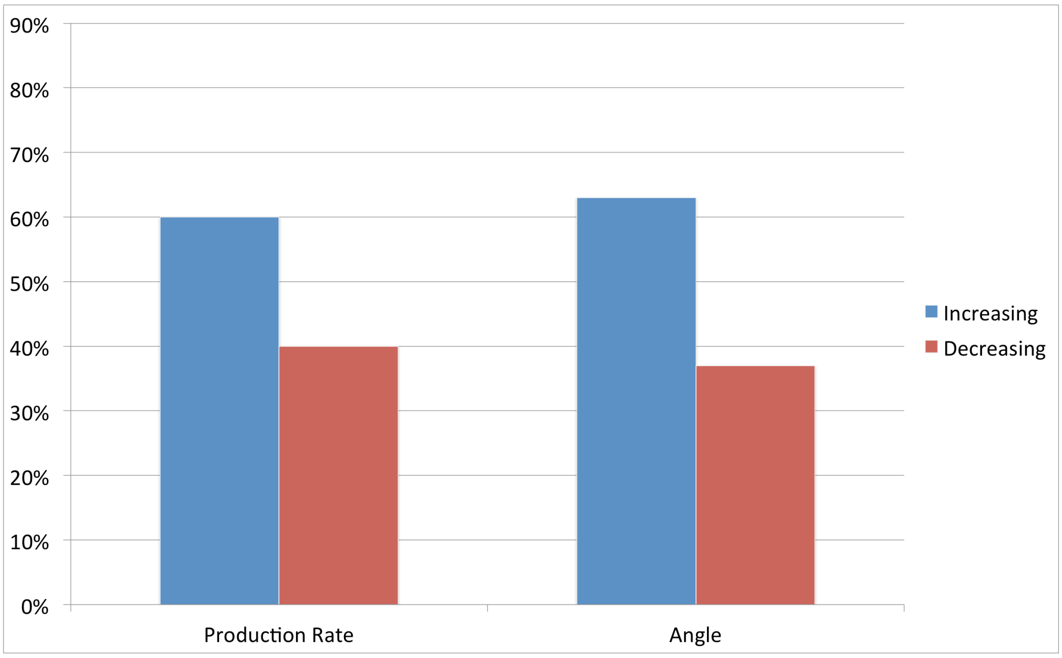


Figure 2.12. Results for Tempo Stimulus C

Tempo stimulus C increased in tempo and the majority of listeners categorised the stimulus as increasing in value for both production rate and angle data as illustrated in Figure 2.12 and recorded in Table 2.3. 60% of listeners categorised the stimulus as increasing for production rate and 63% categorised it as increasing for angle. These numbers are relatively small in comparison to the other results obtained in this evaluation. The McNemar test results for tempo stimulus C, showed that the number of people who interpreted the stimulus differently when the data-type was changed from production rate to angle did not differ to a statistically significant degree X^2 (1, N = 112)=0.42, $p>.05$, phi = .06, the odds ratio is .8. This suggests that the majority of listeners interpret production rate and angle using a positive-polarity mapping and that the number of listeners who interpret production rate with a negative-polarity mapping is not of substantial practical difference to the number of listeners who interpret angle with a negative polarity-mapping. Large minorities of listeners interpret both production rate and angle with a negative-polarity mapping in comparison to the other data collected in this evaluation.

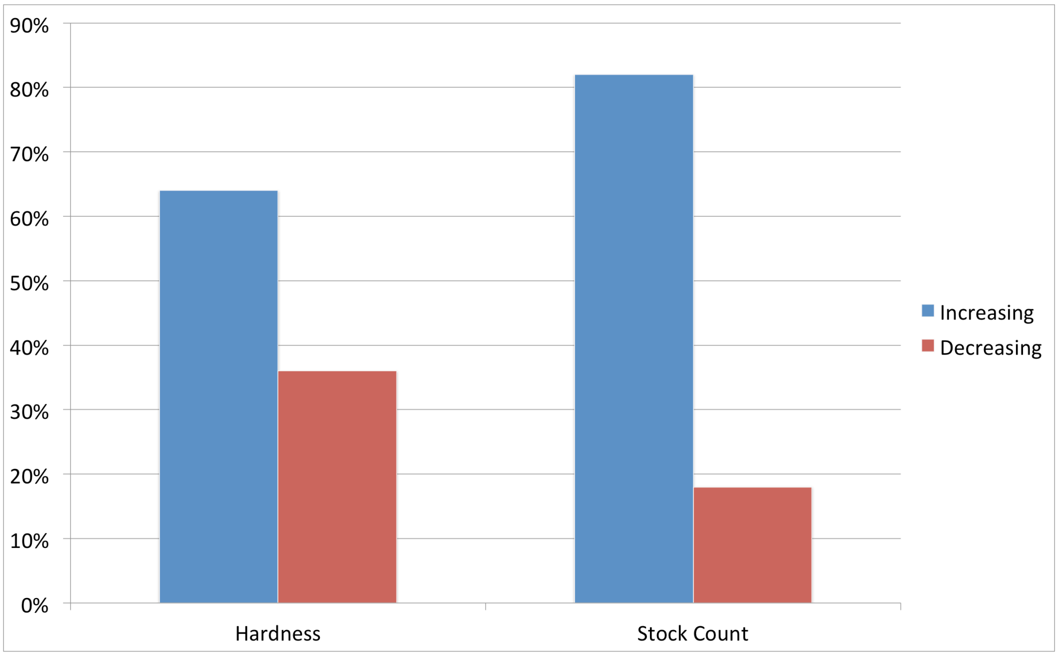


Figure 2.13. Results for Tempo Stimulus D

Tempo stimulus D increased in tempo and the majority of listeners categorised the stimulus as increasing in value for both hardness and stock count data as illustrated in Figure 2.13 and recorded in Table 2.3. 64% of listeners categorised the stimulus as increasing for hardness. This is a relatively small number compared to the other results. The McNemar test results for tempo stimulus D showed that a statistically significant number of people interpreted the stimulus differently when the data-type was changed from hardness to stock count X^2 (1, N = 112)=9.52, $p<.01$, phi = .29, the odds ratio is .35. The effect size represented by phi is medium. This suggests that the majority of listeners interpret hardness and stock count a positive-polarity mapping but that there is a medium sized tendency amongst listeners to interpret hardness with a negative-polarity mapping in comparison to stock count.

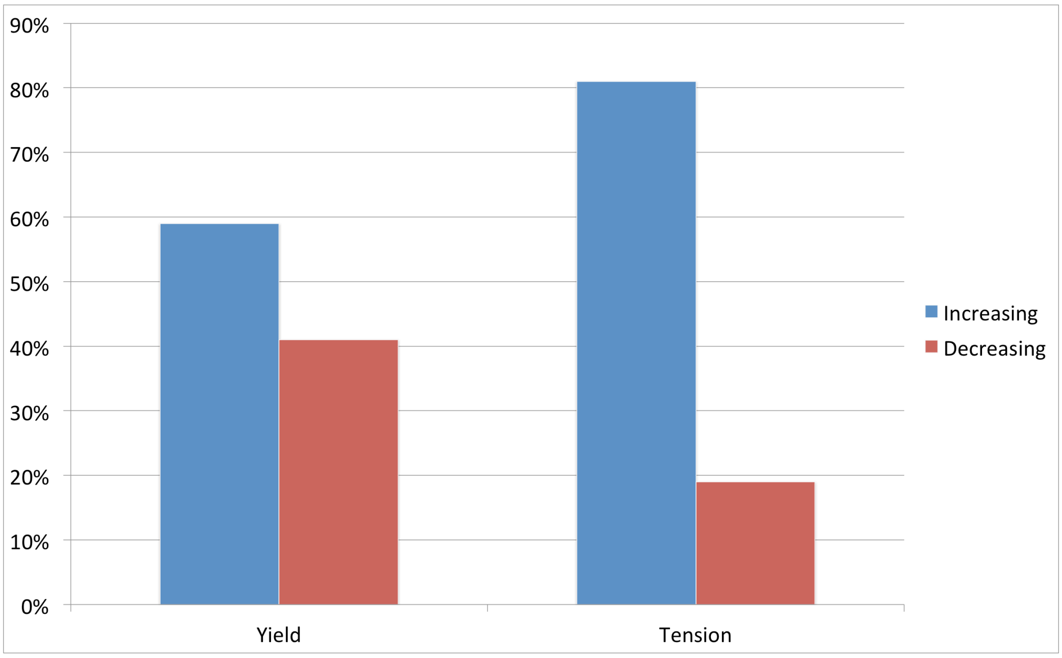


Figure 2.14. Results for Tempo Stimulus E

Tempo stimulus E increased in tempo and the majority of listeners categorised the stimulus as increasing in value for both crop yield and tension data as illustrated in Figure 2.14 and recorded in Table 2.3. 59% of listeners categorised the stimulus as increasing for yield. This is a relatively low figure in comparison to the other results. The McNemar test results for tempo stimulus E showed that a statistically significant number of people interpreted the stimulus differently when the data-type was changed from crop yield to a measurement of physical tension X^2 (1, N = 112)=15.24, $p<.001$, phi = .36, the odds ratio is 4.12. The effect size represented by phi is medium. This suggests that the majority of listeners interpret yield and tension with a positive-polarity mapping but that there is a medium sized tendency amongst listeners to interpret yield with a negative-polarity mapping in comparison to tension.

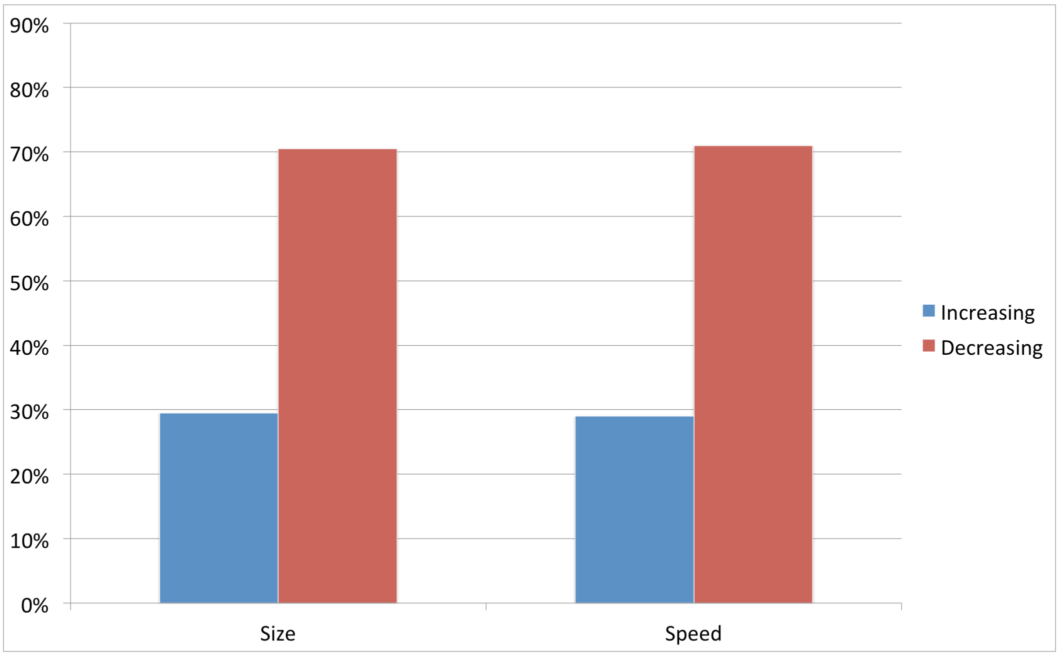


Figure 2.15. Results for Tempo Stimulus F

Tempo stimulus F decreased in tempo and the majority of listeners categorised the stimulus as decreasing in value for both size and speed data as illustrated in Figure 2.15 and recorded in Table 2.3. The McNemar test results for tempo stimulus F showed that a statistically significant number of people interpreted the stimulus differently when the data-type was changed from size to speed X^2 (1, N = 112)=3, $p>.05$, phi = .16, the odds ratio is .33. This suggests that the majority of listeners interpret size and speed using a positive-polarity mapping and that the number of listeners who interpret size with a negative-polarity mapping is not of substantial practical difference to the number of listeners who interpret speed with a negative polarity-mapping.

2.7 Discussion

Positive-Polarity	**Positive w/Negative-Polarity Tendency**
Number of Students	Thickness
Number of Ants	Volume
Concentration	Strength
Population	Stiffness
Absorbency	Hardness
Stock Count	Yield
Tension	Mass
Acceleration	Empty-Space
Temperature	Angle
RPM	Depth
Speed	Production Rate
Size	
Surface Area	

Table 2.4. *Resulting Interpretation Strategies per Data Type.*

The results show that the majority of listeners employed the positive-polarity strategy when interpreting most of the data types. Depth was an exception in which roughly half of the listeners interpreted using a negative-polarity mapping, and half interpreted using a positive-polarity mapping. This result is made more interesting by the fact that the evaluation used a decreasing tempo stimulus rather than a decreasing pitch stimulus would arguably have been more easily interpreted as increasing depth due to cultural conventions linking pitch and vertical height (see Zbikowski, 2005; Neuhoff et al, 2002). Listeners seem to be applying both mapping strategies with equal likelihood to interpret depth. It is here suggested that this is because listeners conceptualise depth in terms of negative physical motion in vertical space, but interpret increasing value in terms of positive physical motion in vertical space.

Listeners were expected to interpret number of students, number of ants, concentration, population, absorbency, stock count, tension, acceleration, temperature, RPM and speed using positive-polarity cognitive mapping strategy. Each of these data types was interpreted with the expected positive-polarity mapping. Listeners also interpreted size and surface area with the positive-polarity cognitive mapping strategy. Listeners were expected to interpret thickness, volume, strength, stiffness, hardness, size, mass, surface area, empty space, depth and angle using negative-polarity cognitive mapping strategies. With the exception of the aforementioned depth result in which roughly 50% interpreted using a negative-polarity mapping, the majority of listeners interpreted these data types using a positive-polarity mapping strategy. Relatively large minorities of listeners interpreted these data types using the negative-polarity mapping, with the exceptions of size and surface area, which were interpreted by clear majorities of listeners using the positive-polarity mapping strategy. The majority of listeners interpreted production rate and yield in a similar manner to these data types, with the majority applying the positive-polarity mapping strategy and large minorities applying the negative-polarity cognitive mapping strategies. Table 2.4 illustrates the cognitive mapping strategies that listeners used to interpret each of the data types.

The results for the tempo stimuli showed a trend whereby a larger number of listeners use a negative-polarity mapping strategy when interpreting tempo stimuli in comparison to the number of listeners who use a negative-polarity mapping strategy when interpreting pitch stimuli. In general the numbers of listeners using negative-polarity mappings were still well below an overall majority for tempo.

These results suggest that the way in which data is described to a listener determines the cognitive mapping strategies that they will use to interpret the sonification to a limited extent and that a listeners might be interpreting sonifications based on their embodied knowledge of the data type being represented. The results also have implications for the hypotheses discussed earlier in this chapter. They provide support for hypotheses A1 and A2 but only supply limited support for hypotheses B1 and B2. The overall listener bias towards positive-polarity mapping strategies suggests that negative-polarity mapping strategies are generally used by minorities of listeners with the exception of depth where roughly half the users apply the negative-polarity mapping strategy. It is here recommended that future research in the area should take this listener bias into account when studying the cognitive mapping strategies employed during sonification listening. These results in turn provide some limited support for the ESLM discussed earlier in this chapter which states that the listeners embodied schematic knowledge of the data being represented, determines the cognitive mapping strategies that they use to interpret a sonification. An early version of the ESLM along with the experiment described here was presented at the International Conference for Auditory Display in Graz, Austria 2015. The paper is available in the conference proceedings (see Roddy and Furlong, 2015b). Though the results generated here are tentative and limited in scope they do suggest that an embodied account of the embodied cognitive aspects of meaning-making in sonification listening could contribute to the design of more communicatively effective sonifications. Additional research is required to further validate the ESLM and the cognitive mapping strategies discussed in this chapter. For example, replication of the experimental results produced here would lend additional support to the model, as would a consideration of different data types and sonic stimuli to those explored here. The evaluations presented in chapters 3, 4 and 5 also examine different components of this model.

2.8 Conclusions

This chapter introduced the Embodied Sonification Listening Model (ESLM) and presented some hypotheses about the cognitive mapping strategies involved in sonification listening on the basis of this model. These hypotheses describe how embodied schematic knowledge might effect a listener's interpretation of a sonification. The chapter then presented an experimental listener evaluation intended to test these hypotheses. The results of these evaluation showed that the majority of listeners tend to apply positive-polarity cognitive mapping strategies for both attribute and number data. However, there is a also strong tendency amongst listeners to apply negative-polarity mapping strategies for sonifications of attribute data. This provides some limited support for the ESLM which claims that the listeners knowledge of the data represented determines how the experience and interpret a sonification and that the cognitive mapping strategies listeners apply during sonification listening are influenced by their embodied schematic knowledge. Although further research is required to validate these preliminary findings suggest that the ESLM may be useful as a guide to sonification research and practice throughout the remainder of this thesis.

Chapter 3: Embodied Sonic Dimensions - Vocal Gestures

3.1 Overview

Chapter 3 explores the embodied sonic dimensions introduced by the Embodied Sonification Listening Model (ESLM). The chapter isolates, defines and empirically evaluates the communicative effectiveness of a number of these dimensions within the domain of vocal gesture. It opens with an analysis of the mapping-problem as it relates to embodied meaning-making in sonification. Next, the chapter introduces and explains the concepts of vocal gestures and stereo space before discussing two prototyping platforms developed to explore the embodied sonic dimensions offered by vocal gestures. A piece of data-driven music entitled *The Human Cost* is then introduced and discussed. The chapter then presents four hypotheses about the embodied sonic dimensions of vocal gestures developed through exploration using the prototyping platforms and during the compositional process for *The Human Cost*. A number of empirical listener evaluations designed to test these hypotheses are then presented and their results are analysed and discussed. The chapter closes with a set of conclusions.

3.2 Embodied Sonification and The Mapping Problem

Chapter 1 isolated parameter mapping sonification (PMSon) as a promising technique that might be used to exploit embodied meaning-making in sonification listening. It also highlights the computationalist meaning-making paradigm that has historically underpinned much of sonification research and practice. Flowers (Flowers, 2005) introduces the mapping problem, which asks how data can be mapped to sound in a way that preserves its information. Worrall (2009) suggests that the mapping problem poses a significant challenge to the effective application of PMSon. This thesis addresses two aspects of the mapping problem. The first is the question of how to create mapping strategies that communicate data effectively to a listener. The second is referred to as dimensional entanglement (Worrall, 2010; 2011; 2013) and is concerned with the intermingling of auditory dimensions traditionally assumed to be separable within the computationalist framework. For example, in PMSon pitch, loudness, duration and timbre are often mapped to unique data (see Grond and Berger, 2011). These dimensions are not independent. They are integrated and changes in one dimension can cause changes in another. This can make it confusing for the listener to interpret a PMSon during sonification listening (see Peres and Lane, 2005; Flowers, 2005; Worrall, 2010; Peres, 2012).

It is here argued that this aspect of the mapping problem is caused by defining sound using dimensions of measurement borrowed from the acoustic and psychoacoustic paradigms. These dimensions divide up sounds along lines that bear little resemblance to a listeners experiences of the original sound. These dimensions are useful for describing and measuring sound in terms of the acoustic waveform, and its perpetual correlates, but they are not necessarily useful dimensions for communicating information in a sonification context. Truax (1984) argues that the prevailing common sense understanding of sound in the West is built around a model of energy transfer. In this model the energy of physical excitations are transferred to physical waveforms that are in turn transferred to sonic experiences in the mind of the listener. He argues that this model is adequate for quantifying sound in terms of physical phenomena but is not sufficient for describing how sound communicates information to a listener. Wishart (1996) makes a similar argument about Western art music. He reasons that as Western art music evolved the focus of composers shifted from creating and organising musical performances to creating and organising written scores. This reduced the rich multi-dimensional spectra of musical discourse to just three primary dimensions: pitch, duration and timbre. These dimensions represent a small sub-set of the many possible dimensions of sonic experience. Worrall (2010) argues that this reductive approach to music is informed by the computationalist theory of mind and that modern music technologies employed to create sonifications are built around this same disembodied framework which fails to account for the role of the embodied performer and the perceptual and cognitive configuration of the embodied listener. The reduction of the rich spectra of sonic experience to non-orthogonal dimensions of pitch, duration, amplitude and timbre, the appropriation of these isolated dimensions as the primary channels for communicating information to a listener and a disregard for the embodied perceptual and cognitive faculties of the listener in interpreting a sonification have all contributed to the mapping problem. New dimensions of sonic communication are required for an embodied approach to sonification that might overcome the mapping problem. For example, it may be more intuitive to sonify rainfall data using the sound of falling rain where the listener would perceive an increase in the heaviness of the rainfall sound as the data level increases. This perceived heaviness of rainfall represents an embodied sonic dimension that cannot be easily described in terms of pitch, amplitude and duration but presents a communicative sonic channel capable of conveying information to a listener nonetheless. The embodied sonic complexes and dimensions defined in Chapter 2 describe sounds as continuous and cohesive perceptual units. They better reflect the listener's embodied sonic experience and it is here argued that they represent useful conceptual models for working with sound in a sonification context. Mapping data to sonic complexes that a user associates with the data source on the basis of their embodied schematic knowledge might addresses the first part of the mapping problem. Restricting those mappings so that each data-stream is associated with a single embodied sonic dimension within a single embodied sonic complex might address the second part of the problem.

3.3 Vocal Gestures

Vocal gestures provide a unique domain of embodied sonic complexes which may be of use to sonification. This thesis defines a vocal gesture as any vocal utterance that communicates meaning through the prosodic, non-linguistic dimensions of speech. The key concept in this description is prosody. Prosody provides a highly evolved, widely familiar, and thoroughly embodied channel for communication (Hirschberg, 2002; Juslin and Laukka, 2003; Armstrong *et al.*, 1995; Grieser and Kuhl, 1988; Grandjean *et al.*, 2005; Elordieta and Prieto 2012; Alba-Ferrara *et al.* 2011) that has much to offer sonification research and practice (Worrall, 2011; Hermann and Ritter, 1999).

Large numbers of people, regardless of gender and culture, share a similar vocal apparatus and a similar auditory system that interprets its utterances. It has been argued that, as a result, human vocal gestures share some systematically stable prosodic properties that are common to across large numbers of people and can act as reliable communicative dimensions, even across cultural and language barriers (Vaissière, 1983). Smalley (Smalley, 1996), in a similar view to Wishart (1996), describes vocal gestures as special cases of physical gestures. He also declares that, when presented in the context of electroacoustic music, vocal gestures grab the listener's attention, compelling them to decode their meanings.

Armstrong *et al.* (1995) provide evidence for the theory that speech evolved as a simulation, or mimesis, of the communicative physical gestures used by pre and early human species. They show that speech is handled in the brain by the same mirror neuronal networks that deal with the more familiar embodied domain of physical gesture indicating that vocal utterances are understood as physical gestures. Gentilucci and Corballis (2006) and Fogassi and Ferrar (2007) provide further empirical neuroscientific evidence for this position. Cox (2001) builds on this thinking in his mimetic theory of music which suggests that musical meaning emerges through the sensorimotor mimesis of the physical gestures required to create musical sounds. Juslin and Laukka (2003) unite these strands of thought in their meta-analysis of emotion in vocal expression which determines the prosodic dimensions of emotional expression common to both musical and speech-based vocalisations.

It is here argued that vocal gestures promise effectively communicative embodied sonic dimensions for sonification because they represent a familiar, intuitively meaningful, domain of shared embodied experience in which rich sonic meanings and the physical gestures of their production are closely entwined. Critically, there is a wealth of empirical neuroscientific evidence attesting to this fact (Armstrong *et al.*, 1995; Armstrong, 2002; Gentilucci and Corballis, 2006; Fogassi and Ferrari, 2007). Additionally a number of auditory display researchers and practitioners have explored and applied vocal gestures in a sonification context. For example, Cassidy *et al.* (2004) used vocal gestures to aid in the diagnosis of hyper spectral colon tissue images while others have applied vocal gestures in the context of EEG sonification (Hermann *et al.* 2006; Baier and Hermann, 2004; Hermann and Baier, 2009, Chafe *et al.*, 2013). Furthermore a number of frameworks for vocal sonification have been developed (see Fox and Carlile, 2005; Grond *et al.,* 2011).

3.4 Stereo Space

Spatial dimensionality is an important feature of the vocal gesture, as all vocal gestures unfold in spatial fields. Stereo audio is a method of sound playback and diffusion that employs two channels of audio and can create a sense of spatial extent for a listener. It can be used to add a sense of spatial dimensionality to synthesised vocal gestures. The work described here focuses on the use of two channel stereo audio diffused through two speakers or headphones, where the channel signals can be further modified to enhance the sense of spatial extent. This is termed ‘stereo-space’ for the remainder of this chapter. Spatial audio techniques have been adapted to develop innovations in the entertainment industry (Rumsey, 2001), the arts (Blesser and Salter, 2009) and in auditory display (Nasir and Roberts, 2007; Groehn and Takala, 1995). This project exploits spatial aspects of auditory cognition for determining embodied sonic dimensions for sonification. Auditory experience has spatial dimensionality, a fact that was explored in Chapter 2. This spatial dimensionality is defined and organised by the listener's embodied cognitive faculties (Cox, 2001; Johnson, 2008; Zbikowski, 2005; Brower, 2000). Space is an integral part of human life with all embodied experiences taking place within spatial fields (Todes, 2001; Johnson, 1987). Embodiment in spatial fields provides a universally familiar domain of human experience within which much of our abstract thinking is grounded (Lakoff and Johnson, 1999). As such listeners make use of these domains of spatial experience to make sense of auditory stimuli (Cox, 2001; Johnson, 2008).

This project also leverages stereo space in order to ensure that the research here described is of effective use across a broad range of hardware platforms. Two channel, two speaker stereo sound has become something of an unofficial standard across the personal computer market for both home and business use and the majority of mobile and smart devices come equipped with stereo playback and a stereo headphone. As such it is suggested here that two channel, two speaker stereo systems are as close to a universal audio standard as can be expected. In focusing on designing solutions for the stereo space paradigm, this project is ensuring that the knowledge produced will be widely applicable across a large range of hardware and playback systems.

3.5 Exploratory Research: Exploration with Prototyping Platforms

Both exploratory and confirmatory research techniques were employed during the research described in this chapter. During the exploratory phase, exploration with two prototyping platforms written in the Csound audio programming language was used to test a set of hypotheses which were derived from a reading of the embodied cognition literature as described in Chapter 1. Hypotheses about the communicative effectiveness of embodied dimensions could be quickly and easily explored by editing, adapting and extending the code for these prototyping platforms. These hypotheses were then tested using empirical listener evaluations in the confirmatory stage.

3.6 Prototyping Platform 1

Both prototyping platforms make extensive use of fof synthesis (*fonction d'onde formantique* or formant wave-function synthesis). Fof synthesis works at the intersection of the time and frequency domains by rapidly generating sine-wave excitations at the millisecond scale in which the frequency level of the excitation determines the centre frequency of the formant area, the production rate of excitations determines the fundamental frequency of the sound and the attack and decay rate of each excitation determine the bandwidth and skirtwidth (at -40db) of the formant area respectively. Multiple formant areas can be added up to create new timbres. Fof[9] synthesis sounds distinctly vocal as it was originally developed to simulate the human voice (see Rodet, Potard and Barriere, 1984). It was extended to Csound by Clarke's fof opcode (1992; 2000) multiple iterations of which can be stacked up to simulate vowel formant profiles. Csound's Spat3d opcode, authored by Istvan Varga (see Vercoe *et al.*, 2015), can position and move input signals through 3D space. It achieves this through the creative use of reverbs, filters and amplitude modulation outputting its audio in a number of useful formats. The opcode proved valuable for exploiting the sense of spatial extent available in two-channel two speaker stereo audio.

The first prototyping platform operated in four stages as illustrated in Figure 3.1. The data processing stage was carried out by instrument 1. Instruments 2, 3 and 4 map the data to synthesised vocal gestures. Instrument 5 spatialises the vocal gestures. Data to be sonified is stored in function tables (f-tables) and read into instr 1, using the tablei opcode at a rate determined by the p3 field in the score. Data values are rescaled here and made globally available for mapping to vocal gestures and spatialisation. This rescaling requires the user to determine the highest and lowest values in the data set and update them to giDatHi and giDatLo, respectively.

Instrument 2 maps this data to modulate between vowel formant profiles determined by p6 and p7 in the score. The vowel formant values are all stored in f-tables below the header and are taken from appendix D of The Canonical Csound Reference Manual (Vercoe *et al.*, 2015). An elaborate series of if and elseif statements is used to select the frequency, amplitude and bandwidth values for the formant profiles. Transform functions are used to change one set of vowel formant values into another at a rate defined by the data. Instrument 3 creates the prosodic and jitter signal for the vocal gesture and maps them to the data. Phrasing, intonation and pausing are simulated using a series of low frequency oscillators (LFOs). Vocal jitter was simulated by mapping the data to generate different kinds of noise.

Instrument 3 contains five fof generator opcodes to which the outputs of instruments 2 and 3 are inserted. The outputs of instrument 4 are passed to instrument 5 for

---

[9] For very high pitched sounds fof synthesis is not effective because listeners cannot properly differentiate the formant profiles. The stimuli in this thesis avoided the use of very high pitched fof sounds.

spatialisation. Instrument 5 feeds the vocal gesture into the spat3d opcode while the data controls the spatialisation parameters to move the vocal gestures through stereo space. This prototyping platform contained a number of extra features which proved to be redundant during the exploratory phase. They are not detailed here but can be seen in the code for the prototyping platform that is provided on the appendices DVD.

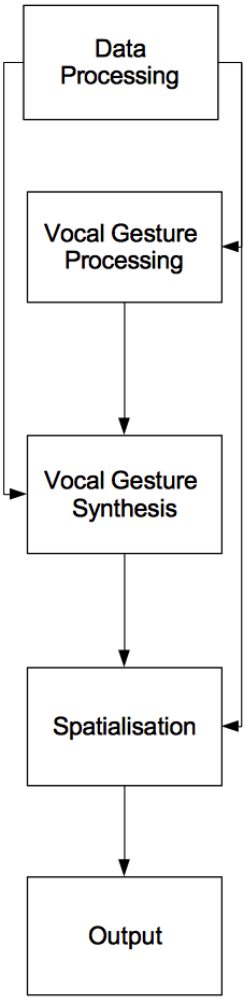


Figure 3.1. Prototyping Platform 1

3.7 Prototyping Platform 2

A second prototyping platform was developed using the Csound audio programming language to build on the functionality and eliminate the redundancies of the previous prototyping platform. The first prototyping platform had a large range of programmable parameters to which data could be quickly and easily mapped within the body of the instrument's code. This was because the useful dimensions of vocal gestures for sonification had yet to be determined. As a result a prototyping tool was needed that would make available for thorough investigation as many different sets of parameters for data mapping as possible. Once these dimensions had been established the generality of the platform could be scaled back in a second platform that could focus on further exploring these dimensions. As a result each individual instrument in the second platform operates as a scaled back version of the entire first prototyping platform.

Certain functionalities deemed no longer relevant to the project have been sacrificed. For example, where the first prototyping platform gave a user direct access to the individual formant parameters of each fof generator, these are locked away in f-tables in this platform because the investigation of the usefulness of these dimensions for sonification, on the previous platform, proved inconclusive. Instead it was determined that vowel sounds provided a good dimension for sonification, so the formant parameters are instead hard-coded to simulate these vowel sounds.

As mentioned the second prototyping platform contains multiple iterations of one single instrument. All instances of the instrument have their own data processing routines allowing each one to process, scale and map data differently as illustrated in Figure 3.2. The data is scaled using the same transform algorithm as the first prototyping platform. Prosody is also generated using the same algorithms and techniques as the first platform. The vowel determination section of this platform differs from the previous one. The large mass of code required to store, access and mix between vowel formant values proved unwieldy and so was slimmed down for this prototyping platform. Isaac Wallis[10] user-defined opcode, vowgen, was extended for this purpose to create vowgen2. Vowgen offered a more economical, in terms of code, and robust system for generating vowel sounds using fof synthesis than that of the first prototyping platform.

Vowgen nests the f-tables for each set of vowel formant values inside another f-table named index. This f-table acts as an index of all possible vowel formant values for a specific voice type, defined by the user in the imode argument of the opcode. The ftmorf opcode is used to transform between entire formant value tables along the index updating the transformed formant values to another f-table, named imorf, in real-time. The formant values of the imorf table are then read into five fof generators which generate the signal for the vowel sounds.

---

[10] http://www.csounds.com/udo/displayOpcode.php?opcode_id=76

Vowgen2 extends the functionality of vowgen by accepting kris, kdur and kdec values which determine the envelopes of each individual fof grain. This useful addition allowed for variation of the timbres across multiple instances of the instrument making it easier to distinguish between individual vocal gestures. The instruments in this prototyping platform used a simpler spatialisation technique than that used in the original platform where complex reverb patterns were computed by simulating room acoustics using the spat3d opcode. A left-right panning system, using the pan2 opcode, coupled with reverb modelling, using the freeverb opcode, was used to move the vocal gestures produced by the vowgen2 opcodes through the stereo-spatial field in keeping with the data changes. Distance can then be modelled by adjusting the mix between the original vocal gesture signal and the reverb signal. Increasing reverb while decreasing the signal is perceived as an increase in distance from the signal to the listener. Decreasing reverb while increasing the signal is perceived as a decrease in distance. Although the spat3d opcode used in the first prototyping platform was somewhat effective its results were often too subtle, making the dynamic spatial profile of the vocal gestures across the X, Y and Z axis, difficult to determine. Also the opcode's capacity for vertical movement of a sound source was rendered redundant as both the literature (Cox, 2001; Zbikowski, 2005) and explorations, carried out with the previous platform, suggested that listeners interpreted pitch in terms of vertical orientation with high pitches mapping to higher vertical positions and low pitches mapping to lower vertical locations. The code for this prototyping platform is included in the appendix for this chapter and is included in the digital appendices for this chapter on the appendices DVD.

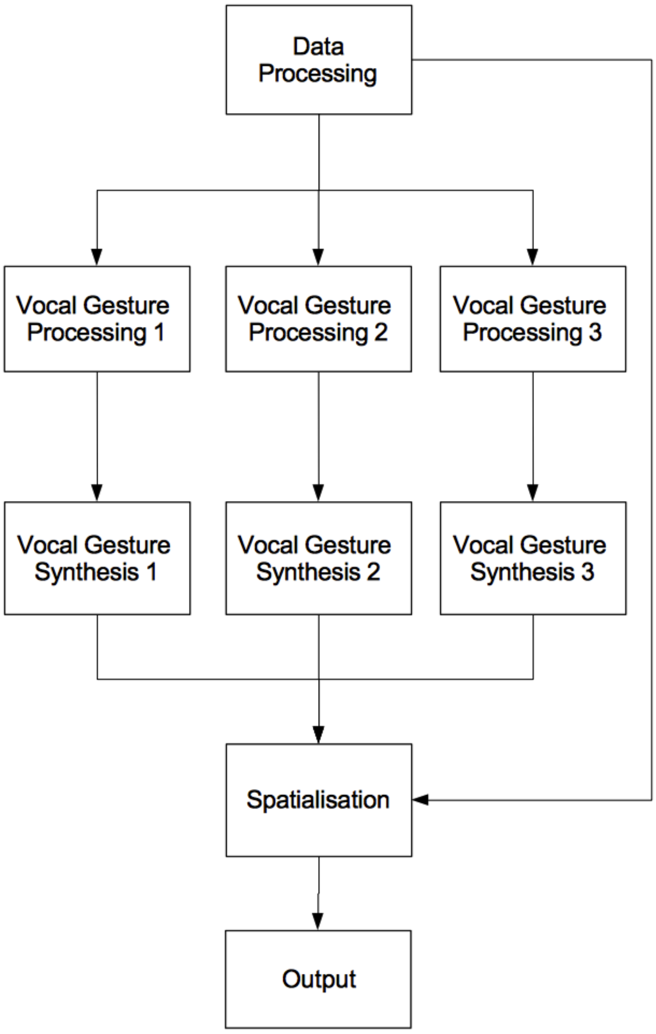


Figure 3.2. Prototyping Platform 2

3.8 Application and Dimensions

The prototyping platforms were used to explore and establish embodied dimensions using the unique dimensions of embodied schemata, tension and release patterns and emotional dimensions offered by vocal gestures. Embodied schemata have been employed in numerous sonification and auditory display applications (Roddy and Furlong, 2013a). A list of useful embodied schemata as they apply to human computer interaction is presented by Macaranas *et al.* (2012). These are inspired by the work of Johnson (1987) and the later work of Hurtienne and Israel (2007) and Evans and Green (2006). The speed, size, multitude, spatial and roughness schemata as they relate to vocal gesture and spatial location and movements have been determined as the specific embodied schematic dimensions of interest in this chapter. The process by which these dimensions were chosen is presented appendix 3.4 of this thesis.

Certain vowel sounds require more tension in the throat and facial musculature of the speaker than others (Durand, 2005). From this it was hypothesised that patterns of tension and release could be generated using vowel sounds. It was thought that this would represent a uniquely embodied approach to meaning-making in sonification as sounds would be identified as either lax or tense on the basis of the physical gestures used to create them in the human vocal tract. Considering to Cox's (2001) mimetic hypothesis, a listener might understand the tension in a vowel sound, by simulating the bodily processes involved in the production of the vowel in the sensorimotor system.

Juslin and Laukka (2003) isolate a number of cues common to both vocal and musical expression of emotion that are generally compatible with Scherer's (1986) vocal affect model, a standard of sorts for the field of emotional psychology. They believe that the same cognitive process underlies emotionally expressive vocal and musical expression, and that musical cognition is an outgrowth of vocal cognition. This is generally in agreement with the EC hypothesis which assumes that all higher level cognition grows out of lower level sensorimotor processes (Varela *et al.*, 1991). This chapter treats these cues as indicative of the embodied sonic dimensions that are inherent in vocal gestures.

3.9 Composition as a Research Method

Chapter 1 isolated the mapping of data to musical structures, as exemplified by earcons, as a sonification method that could be used to exploit the embodied aspects of meaning-making. Historically, music and computer music have been critical contributing fields to sonification research and practice. An increasing number of researchers in the area place the role of aesthetics as a primary concern and call for a more focused investigation of the aesthetic dimensions of sonification (Grond and Hermann, 2012; Roddy and Furlong, 2014; Barrass and Vickers, 2011, Vickers and Hogg, 2006). This thesis focuses on cognitive meaning-making in sonification. Aesthetics and cognitive meaning-making are closely intertwined and it is the listeners' cognitive meaning-making faculties that drive both aesthetic and mundane experience alike (see Johnson, 1987; 2008). In this point of view, aesthetic experiences are those which effectively exploit a person's embodied cognitive faculties to create particularly meaningful experiences. A similar line of thought is developed by Fauconnier and Turner (2002) and is extended to music and auditory cognition by thinkers like Zbikowski, Brower, Larsson, Cox and Kendall as discussed in Chapter 1. As such, a fully realised and serious account of the role of the human body in sonification listening, a practice chiefly concerned with making data meaningful, must necessarily account for aesthetic experience. This would suggest that when a sonification is effectively communicating data to a listener, it is creating a more heightened aesthetic experience for that listener, as discussed in chapter 1. This sentiment is echoed by Vickers and Hogg (2006) who argue that it is not necessarily useful or meaningful to distinguish between musical and non-musical sonifications. On this basis, research via composition was chosen as a technique for exploring and uncovering useful communicative dimensions for sonification.

Throughout the course of the research presented in this thesis three data-driven compositions were completed. These are *The Human Cost, Idle Hands* and *Doom & Gloom.* They are described in detail in appendices of this thesis. The compositional process for each piece provided an opportunity for exploratory research. As mentioned previously hypotheses about embodied sonic dimensions and their relevance to sonification practices were developed using the prototyping platform and were also developed through the compositional process. These hypotheses could then be tested through empirical listener evaluation.

3.10 Composition: *The Human Cost*

The Human Cost is a data driven music composition that was developed while conducting exploratory research, which aimed to find communicatively effective embodied dimensions for sonification. The compositional process for this piece is described in depth in appendix 3.4. The technical aspects of the piece are discussed here. The composition of the piece had two aims. The first was to guide the exploration into embodied sonic dimensions and the second was to explore the aesthetic possibilities offered by embodied sonic dimensions. *The Human Cost* maps data to sonic complexes and dimensions provided by the domain of vocal gestures. The piece expresses the human cost of Ireland’s economic crash. There were three phases to its composition. The first two adapted the second prototyping platform discussed previously to map data to the embodied sonic dimensions of the vocal stereo spatial embodied complex. The third phase consisted of additional compositional, editing and mixing work required to organise those vocal dimensions into a cohesive whole. The second prototyping discussed previously was used to generate sonified material for this composition. The piece mapped data to the dimensions of tension, size and spatial location in vocal gesture. The composition was intended to furnish listeners with an understanding of how Ireland’s recent economic recession has impacted its citizens. It was not intended to accurately convey individual data-points.

The Deprivation Rate, Unemployment Rate and Emigration Rate from 2007 to 2012 were used to drive three fof synthesis algorithms based around the extended vowgen2 opcode. As each of these economic indicators increased, the perceived tension in the human vocal simulations would increase also. This is achieved by a shift in formant shapes from an ‘A’ shape to an ‘E’ shape. The rationale behind this design choice is that ‘E’ sounds require a tightening of the musculature in the throat around the vocal apparatus and ‘A’ sounds require a loosening of these muscles. As a result it is hypothesised that listeners might perceive an ‘A’ sound as relaxed and an ‘E’ sound as tense.

The outputs of each of these algorithms retained a harmonic relationship. Algorithms 1 and 2 express the Deprivation and Unemployment rates respectively while algorithm 3 expresses the Emigration rate. The first two algorithms provide a harmonic backing and are therefore devoid of prosody. The third algorithm is differentiated from the other two by way of having strong prosodic variations intended to evoke the sense of a lament. It acts as a focal point akin to a lead instrument in more traditional musical forms.

The panning algorithm is also used to adjust the size of the stereo image for the third vocal gesture narrowing the image for mid level data values and expanding it for both the higher and lower extremes of the data. The vocal gestures were organised in stereo space using the prototyping platform’s panning algorithm. In keeping with the data, the vocal gestures of algorithms 1 and 2 move from right to left.

A fourth algorithm utilises a heartbeat metaphor, in that it sounds like a beating heart, which is mapped to express the fall in GNP between 2007 and 2012. This adds a rhythmic grounding to the piece. The algorithmic processing was undertaken in Csound. These algorithms output sonifications as .wav outputs which were then loaded into Logic Pro X. Here some reverb and delay modelling were applied. In keeping with the overall aesthetic theme these delay and reverb effects were controlled by the data. The second prototyping platform was further extended in Csound to map the data to MIDI control change values and send them to Logic Pro X where they were assigned to control the overall mix of the piece. Once the effects were applied in this manner, they were tuned and reorganised by hand to make for a more balanced and effective composition.

The composition is included on the accompanying appendices DVD. *The Human Cost* was chosen for performance during the Contemporary Music Centre's Spring Salon Series at the Kevin Barry Room of Ireland's National Concert Hall in 2015. This was the first salon session by the CMC to focus solely on electronic music and the piece was very well received by its audience. Listeners' interpretation of the aesthetic merits of the piece were not empirically evaluated. The piece is intended to illustrate how an embodied cognitive approach to sonification can inform the composition of data-driven music.

3.11 Exploratory and Compositional Research Results

Exploration with the prototyping platforms and research via composition documented in appendix 3.4 resulted in the following hypotheses:

(1) Listeners perceive embodied sonic dimensions of size, speed, amount and texture by applying the relevant embodied schematic organisation patterns to synthesised vocal gestures.

(2) Listeners perceive dimensions of tension, weight, strength, heat, size and brightness on the basis of vowel formant profiles in synthesised vocal gestures.

(3) Listeners perceive dimensions of happiness, anger, fear and sadness in synthesised vocal gestures on the basis of Juslin and Laukka's (2003) cross-modal patterns of acoustic cues for discrete emotions.

(4) Listeners perceive dimensions of proximity, vertical orientation, horizontal orientation, approach-passage-retreat in synthesised vocal gestures on the basis of spatial cues.

A number of experimental listener evaluations designed to test these hypotheses are presented shortly. These evaluations were designed to determine the extent to which listeners perceive the dimensions in question on the basis of the synthesis parameters used.

3.12 Synthesising Vocal Gestures

Vocal gestures can be synthesised using numerous computational audio synthesis techniques. The techniques used for the empirical testing phase of this research differed from those used to develop the prototyping platforms discussed earlier. The decision was made to use Native Instrument's Reaktor 5 sound design platform, running the Razor additive synthesis engine, to generate vocal gestures, and Izotope's Ozone 5 Mastering suite to generate stereo spatial movements. Both were run as plugins in Logic Pro X. Razor allowed for additive synthesis driven by a bank of parallel sine oscillators generating 320 individual partials. The amplitude of each individual partial can then be augmented directly. For this project, the partials were configured to synthesise vocal gestures. Ozone allowed for realistic reverb modelling using a variety of room impulse responses that, in tandem with Logic Pro X's imaging plugins were useful for generating stereo spatial movement. This approach was selected for creating stimuli because these technologies presented a much more sophisticated means of generating vocal gestures and stereo spatial movements than those afforded by the prototyping platforms. The audible results they produced were superior to those of the prototyping platforms. Csound's spat3d opcode often produced overly subtle results making it hard for the listener to interpret. Freeverb often produced overly coarse results where the spatialisation sounded unrealistic and cartoon like. Ozone 5's reverb modelling algorithms produced more convincing results. This was likely due to the use of high quality impulse responses to model spaces. Although Ozone 5 does not allow the user to modulate as many parameters of the sound as Csound, the parameters it does offer are very well tuned and refined. Csound's fof opcode often produced unconvincing vocal simulations and indistinct vowel sounds. Reaktor produced extremely realistic vocal timbres and clear vowel sounds. This may be a result of the additive synthesis techniques used and Reaktor's complex synthesis engine, which has been designed for commercial applications and actively refined and expanded since 1996.

3.13 Evaluation Recruitment Method

Participants were recruited through the online crowdsourcing platform Crowd Flower.[11] Each evaluation was designed, hosted and delivered on the Survey Gizmo[12] web-platform. Precautionary measures were taken to ensure that participants were using

---

[11] www.crowdflower.com

[12] www.surveygizmo.com

proper equipment. The measures taken are discussed in greater detail in section 2.4.1. All participants were required to pass a validation test to prove that they were undertaking the evaluations using a 2-channel stereo setup with either a good set of headphones or a 2 speaker array. Potential participants who did not pass the validation test were not allowed to take part in the evaluations. Participants were recruited from a large international pool of 41 countries to ensure that the results obtained were not specific to a particular culture but could be generalised across a large and varied selection of people. Participants were financially compensated for their participation according to standard crowdflower rates. 139 participants took part in the evaluations. Of that number 26% were female and 74% were male. 20% of listeners had formal musical training and 27% played an instrument. Additional demographics information is supplied in the appendices for this chapter.

Listeners completed a short training trial to familiarise themselves with the procedure at the outset of the evaluation. The evaluation took roughly 45 minutes to complete. The parameters discussed in this chapter are relative rather than absolute. For example, a fast stimulus will only be perceived as fast in relation to a slow stimulus. More details on all of the synthesis parameters used to create the stimuli in each of these evaluations, and some sample evaluation questions from each evaluation are included in the appendices for this chapter. The stimuli themselves, the entire set of evaluation questions, and all the raw data collected during these evaluations, are included in the digital appendices for this chapter on the appendices DVD. Listeners undertook the each evaluation in a set order and as such it is possible that there were ordering effects. These effects would be unlikely to determine the results, due to the relative nature of the judgements investigated in this evaluation.

3.14 Evaluating Hypothesis (1)

The following four evaluations were designed to test hypothesis (1) which holds that listeners perceive embodied sonic dimensions of size, speed, amount and texture by applying the relevant embodied schematic organisation patterns to synthesised vocal gestures. Relevant patterns for organising synthesis parameters of vocal gestures were uncovered during exploration using the prototyping platforms and research via composition. These parameters can be used to model and control these embodied sonic dimensions in vocal gestures. Each of the evaluations discussed in this chapter used the minimum number of stimuli to generate a reliable result. The reasoning for the number of stimuli used in each evaluation is provided in the description of design and materials for each evaluation.

3.15 Hypothesis (1) Evaluation A: Speed

Synthesis Parameters	Fast Pole	Slow Pole
Number of Events	More	Less
Length of Events	Shorter	Longer
Pauses Between Events	Shorter	Longer
Speed of Vowel Changes	Faster	Slower
Event Attacks	Shorter	Longer
Event Decays	Shorter	Longer

Table 3.1. *Vocal Gestures: Relative Speed Parameters*

This experiment is intended to determine if listeners perceive changes in the speed dimension of a synthesised vocal gesture and to establish the validity of the synthesis parameters, derived through exploration with the prototyping platforms and compositional research, used to model this embodied sonic dimension. Johnson (1987) suggests that subjects develop a Fast-Slow schema through frequent embodied encounters with processes and entities that change in speed in one’s environment. As such, listeners should be able to identify changes in speed along a relative bi-polar scale from fast to slow on the basis of this embodied schema. Research via composition and explorations with the prototyping platforms implied that in order for one vocal gesture to be considered faster than another it must contain more discrete events (for example words spoken), these events must be of a shorter duration, there must be shorter pauses between each event, the vowel formant profile must change at a fast pace and the gesture must contain quick attack and decay times. These parameters are listed in Table 3.1. The process by which these parameters were discovered and the context in which they are effective is described in appendix 3.4.

3.16 Design and Materials

Participants listened to 12 stimuli of differing lengths from 10 to 26 seconds. 6 stimuli were clearly pitched and 6 had noisy timbres in order to obtain results that hold for both clearly pitched and noisy vocal gestures. Each set of 6 stimuli was further subdivided into 2 groups of 3, a fast group and a slow group. One stimulus had all of the synthesis parameters to model the fast pole, and another had all of the synthesis parameters to model the slow pole of the hypothesised speed dimension. Another two stimuli have half of the parameters for both fast and slow respectively and the last two stimuli have random parameters for both the fast and slow poles. This design allowed for listeners ratings of the stimuli with all of the parameters applied to be compared with their ratings of stimuli with only half of the parameters applied and stimuli with random parameters applied for both fast and slow and noisy and pitched stimuli. The exact parameters are presented in appendix 3.5. After taking part in a short training exercise to familiarise themselves with the evaluation procedure, listeners were presented with each of the stimuli and asked to rate them on a five point Likert scale consisting of Very Slow (1), Slow (2), Medium (3), Fast (4) and Very Fast (5). They were instructed to listen to the sounds as many times as needed to help rate the stimuli.

3.17 Results and Analysis

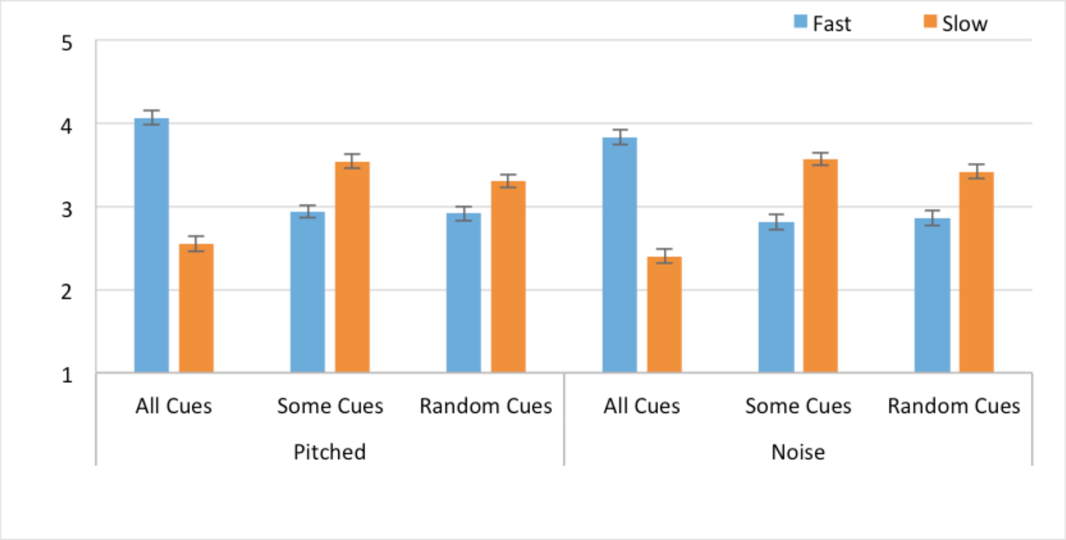


Figure 3.3. Ratings of speed stimuli. Error bars are standard error of the mean

The results for this evaluation are illustrated in Figure 3.3. Tables 3.2 and 3.3 show the mean results for Likert ratings of fast and slow parameters for both pitch and noise stimuli respectively. The tables show standard deviations while Figure 3.3 shows standard error. It can be seen from the tables that when all of the parameters were applied to the stimuli the listeners' ratings improved. They rated the stimuli that had all of the fast parameters applied to be fastest and the stimuli with all of the slow parameters applied to be slowest for both pitch and noise. However, the mean rating for the pitch stimulus with the fastest parameters (4.06) was closer to a fast rating (4) than a very fast rating (5) and almost fell (.06) within one standard deviation (1) of the midpoint of the scale (3). The noise stimulus with the fastest parameters (3.83) was rated more closely to fast than to very fast and fell within one standard deviation (1.06) of the midpoint of the scale. The mean rating for the pitch stimulus with all the slow parameters applied (2.55) falls within one standard deviation (1.07) of the midpoint and is slightly closer (.45) to the midpoint than it is to the rating for a slow speed (2). The noise stimulus with all the slow parameters applied (2.4) was rated more closely to slow (2) than very slow (1) and fell within one standard deviation (1) of the midpoint. These initial results might be ambiguous as the standard deviations show that the mean ratings could be accounted for by chance. They also show a bias whereby listeners tend towards rating all the stimuli as fast and pitched stimuli were rated faster than noise stimuli. The slowest rated clip is only .6 of a Likert category slower than medium while the fastest is rated 1.06 faster than medium.

Pitch Stimuli	**Mean**	**Std. Dev.**
All Fast	4.06	1
Some Fast	2.94	.91
Random Fast	2.91	.99
All Slow	2.55	1.07
Some Slow	3.54	1
Random Slow	3.3	.91

Table 3.2. *Mean Likert Rating Results for Pitch Stimuli*

Noise Stimuli	**Mean**	**Std. Dev.**
All Fast	3.83	1.06
Some Fast	2.81	1.08
Random Fast	2.86	1.09
All Slow	2.4	1
Some Slow	3.57	0.91
Random Slow	3.41	1

Table 3.3. *Mean Likert Rating Results for Noise Stimuli*

A repeated measures ANOVA was performed on the speed measures with a 2 (tone: pitch vs. noise) x 2 (parameter type: fast vs. slow) x 3 (parameter level: all vs. some vs. random) design. There was no main effect of tone F(1, 138)=2.54, p>.05, $\eta_p^2 = .01$, parameter type F(1, 138)=2.89, $p > .05$, $\eta_p^2 = .02$, or parameter level F(2, 276)=1.91, $p > .05$, $\eta_p^2 = .01$. No main effect was expected here, as the main point of interest was in any interactions between parameter type and level.

There was an interaction between tone and parameter level F(2, 276)=3.37, $p < .05$, $\eta_p^2 = .024$. Paired contrasts were performed to decompose the interaction. They showed that there was a statistically significant effect of parameter level on pitched stimuli F(2, 137)=4.78, $p < .01$, $\eta_p^2 = .065$. The small effect size indicated a small degree of practical significance. There was no significant effect of parameter level on noise stimuli F<1. In practical terms, this suggests a bias whereby the application any level of speed parameter to a pitched stimulus results in the listener interpreting it as faster to a small degree. However the same does not apply for noise stimuli.

Listeners rated pitched stimuli faster than noise stimuli when all parameters were applied F(1, 138)=7.52, $p < .01$, $\eta_p^2 = .05$, but did not rate pitched stimuli faster than noise when only some parameters F<1, or random parameters F<1 were applied.

This suggests a trend (on the basis of *p* value and medium effect size) whereby listeners interpreted fast pitch stimuli as faster than fast noise stimuli and slow pitch stimuli as faster than slow noise stimuli to a medium degree when all parameters were applied.

There was also an interaction between parameter level and parameter type F(2, 138)=134.8, $p < .001$, $\eta_p^2 = .49$. Paired contrasts performed to decompose the interaction showed that stimuli with all the slow parameters applied were judged to be less fast than stimuli with some or random slow parameters applied F(2, 137)=58.09, $p < .001$, $\eta_p^2 = .46$. Likewise, stimuli were judged to be faster when all fast parameters were applied than when some or random fast parameters were applied F(2, 137)=64.56, $p < .001$, $\eta_p^2 = .49$. The significant *p* values and large effect sizes for these analyses suggest results of practical significance. They indicate a strong association between the application of fast and slow parameters and the listeners' perception of fast and slow in vocal gestures.

The analysis also showed that stimuli with all fast parameters were judged to be faster than stimuli with all slow parameters to a large degree of practical significance, F(1, 138)=156.16, $p < .001$, $\eta_p^2 = .53$. However, stimuli with some slow parameters were judged to be faster than stimuli with some fast parameters F(1, 138)=45.24, $p < .001$, $\eta_p^2 = .25$, and stimuli with random slow parameters were judged to be faster than stimuli with random fast parameters F(1, 138)=26.72, p<.001, $\eta_p^2 = .16$. These three variables did not interact, F<1.

These results indicate that the application of all fast parameters results in the listeners perceiving a stimulus as faster than a stimulus with all of the slow parameters applied. They also indicate that the application of all slow parameters result in the listeners perceiving a stimulus as slower than a stimulus with all of the slow parameters applied. As anticipated, the application of some or random parameters result in listeners misinterpreting the stimuli. The results also suggest a bias towards interpreting pitched stimuli, but not noise stimuli, as fast. This is especially true when all fast or all slow parameters were applied. Overall the results suggest that for a large audience of non-specific sonification listeners the parameters explored here are of moderate use in controlling the listeners' perception of speed in a vocal gesture.

3.18 Hypothesis (1) Evaluation B: Size

This experiment is intended to determine if listeners perceive changes in the size dimension of a synthesised vocal gesture and to establish the validity of the synthesis parameters, derived through exploration with the prototyping platforms and compositional research, used to model this embodied sonic dimension. Johnson (1987) suggests that subjects develop a Big-Small schema through frequent embodied encounters with processes and entities that change in size in one's environment. As such, listeners should be able to identify changes in size along a relative bi-polar scale from big to small on the basis of this embodied schema. The relative synthesis parameters, found to be effective for modelling this dimension were determined using prototyping platforms and compositional research. These are listed in Table 3.4. The process by which these parameters were discovered is documented in appendix 3.4.

3.19 Design and Materials

Participants listened to 8 stimuli of 16 seconds length each. 4 of the stimuli were clearly pitched while the other 4 had a noisy timbre. 4 stimuli, 2 with clear timbres and 2 with noisy timbres had the hypothesised parameters for a small vocal gesture applied. Of these 4 stimuli 1 clear stimulus and 1 noisy stimulus were pitched to an A5 and another clear stimulus and noisy stimulus were pitched to an A1. In this way the relationship between pitch and the perception of small vocal gestures could be examined by comparing the results for the small A5 stimulus to the small A1 stimulus.

Another group of 4 stimuli, 2 with clear timbres and 2 with noisy timbres had the hypothesised parameters for a big vocal gesture applied. Of these 4 stimuli 1 clear stimulus and 1 noisy stimulus were pitched to an A0 and another clear stimulus and noisy stimulus were pitched to an A1 In this way the relationship between pitch and the perception of big vocal gestures could be examined by comparing the results for the big A0 stimulus to the big A1 stimulus. After taking part in a short training exercise to familiarise themselves with the evaluation procedure, listeners were asked to listen to each of the stimuli and to rate each one on a five point Likert scale of Very Small (1), Small (2), Medium (3), Big (4) and Very Big (5). Listeners were instructed to listen to the sounds as many times as needed to help rate the stimuli.

Sonic Parameters	**Big**	**Small**
Amplitude	Higher	Lower
Energy Profile	Less High Frequency Energy	More High Frequency Energy
Pitch Level	Lower	Higher
Attack Speed	Slower	Faster
Vowel Profile	A	I
Reverb Amount	More	Less
Dynamics Range	Small	Bigger
Stereo Image	Wider	Narrower

Table 3.4. *Vocal Gestures: Relative Size Parameters*

3.20 Results and Analysis

The results for this evaluation are illustrated in Figure 3.4. Tables 3.5 and 3.6 show the mean results and standard deviations for Likert ratings of big and small parameters for pitch and noise stimuli respectively along with standard deviations. All of the means are clustered around the mid-point of the Likert scale and they all fall within one standard deviation of it, indicating a smaller dynamic range. This suggests that listeners had trouble interpreting the stimuli. Listeners did not rate the size each of the stimuli in the expected order. The Big A1 pitch (3.4) and noise (3.37) stimuli were rated as roughly the same size as the Big A0 pitch (3.34) and noise (3.35) stimuli. The Small A1 pitch (3.04) and noise (3.19) stimuli were also rated as roughly medium (3).

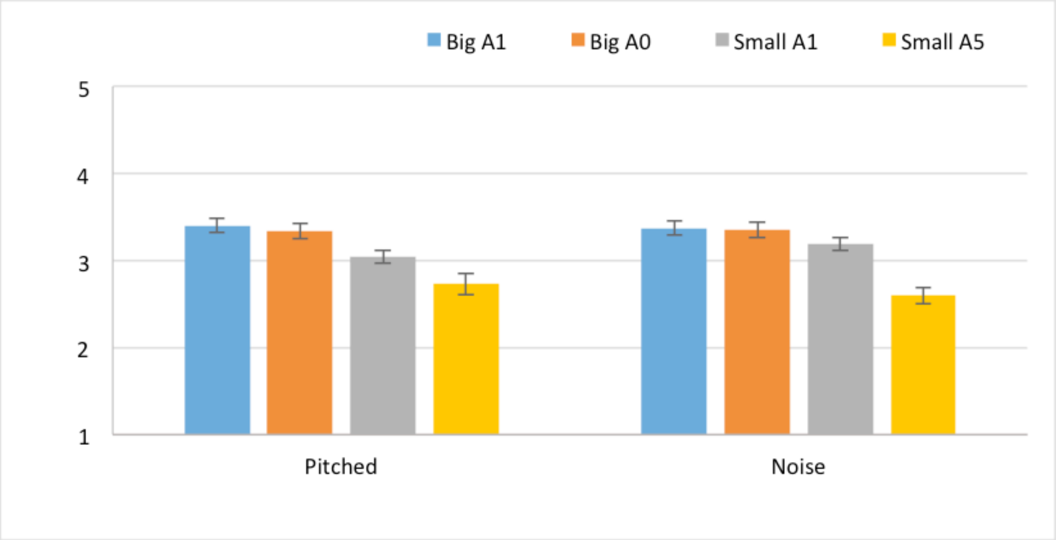


Figure 3.4. Ratings of size stimuli. Error bars are standard error of the mean

The Big A1 pitch (3.4) and noise (3.37) stimuli were rated as roughly the same size as the Big A0 pitch (3.34) and noise (3.35) stimuli. The Small A1 pitch (3.04) and noise (3.19) stimuli were also rated as roughly medium (3).

Noise Stimuli	**Mean Likert Ratings**	**Std. Dev.**
Big A0	3.35	1.09
Big A1	3.37	0.94
Small A1	3.19	0.95
Small A5	2.6	1.13

Table 3.5. *Mean Likert Scale Ratings and Standard Deviations for Noise Stimuli*

Pitch Stimuli	**Mean Likert Ratings**	**Std. Dev.**
Big A0	3.34	1.04
Big A1	3.4	0.94
Small A1	3.04	0.83
Small A5	2.73	1.43

Table 3.6. *Mean Likert Scale Ratings and Standard Deviations for Pitch Stimuli*

The results were further analysed with a repeated measures ANOVA which was performed on the size measures with the design 2(tone: pitch vs. noise) x 4(parameters: Big A0 vs. Big A1 vs. Small A1 vs. Small A5). There was no main effect of tone F<1.

This suggests that listeners interpreted pitched and noise stimuli as equal in size. There was a significant effect of parameters F(3, 414)=19.77, $p<.001$, $\eta_p^2 = .125$, suggesting that the application of size parameters affected the listeners perception of a stimuli's size so that listeners interpreted the big stimuli as much larger than the small stimuli, on the basis of *p* value and effect size. Tone and parameters did not interact F(3, 414)=1.56, $p>.05$ $\eta_p^2 = .011$, suggesting that the application of size parameters affected listeners interpretation of noise and pitched stimuli equally.

It could be argued that pitch played a role in the perception of size for this evaluation. A 4 octave increase in pitch from A1 to A5 for the small stimuli correlated to the perception of decreasing size but a decrease of 1 octave in pitch from A1 to A0 for the big stimuli did not correlate to a decrease in size. Overall the results suggest that the parameters explored here are of limited use in controlling the perception of size in a vocal gesture.

3.21 Hypothesis (1) Evaluation C: Amount

This experiment is intended to determine if listeners perceive changes in the amount dimension of a synthesised vocal gesture and to establish the validity of the synthesis parameters, derived through exploration with the prototyping platforms and compositional research, for mapping data to those dimensions. Changes in the amount dimension happen along a relative bi-polar scale that runs from big to small. The synthesis dimensions obtained through exploration with the prototyping platforms and compositional research are described in detail in the appendix for this chapter.

Johnson (1987) suggests that subjects develop a Many-Few schema through frequent embodied encounters with processes and entities that change in amount in one's environment. As such, listeners should be able to identify changes in amount along a relative bi-polar scale from many to few on the basis of this embodied schema. The relative synthesis parameters, found to be effective for modelling this dimension were determined using prototyping platforms and compositional research. The process by which these parameters were discovered is documented in appendix 3.4.

3.22 Design and Materials

This test compared two approaches to synthesising the amount dimension in vocal gesture. It aimed to determine whether chorusing techniques or a simple layering up of multiple vocal gestures was more effective at modelling the amount dimension.

Participants listened to 4 stimuli of 10 seconds length each. One of the stimuli contained a large amount of vocal gesture tracks and another contained very few. The addition of multiple tracks to a piece of audio is a technique referred to as dubbing. Two more stimuli were created using the chorusing technique. Chorusing is a signal processing technique often used in music production where an input signal is copied multiple times and then a parameter, usually the pitch, is modulated by an LFO. When this slightly altered signal is added back to the original signal it creates the impression of multiple versions of the original sound playing in unison. One stimulus, the stimulus intended to correspond to small amounts, had no chorusing applied while another, intended to correspond to large amounts, did have chorusing applied.

After taking part in a short training exercise to familiarise themselves with the evaluation procedure, listeners were asked to listen to each of the stimuli and to rate each one on a five point Likert scale of Very Few (1), Few (2), Medium (3), Many (4) and Very Many (5). They were instructed to listen to the sounds as many times as needed to help rate the stimuli.

3.23 Results and Analysis

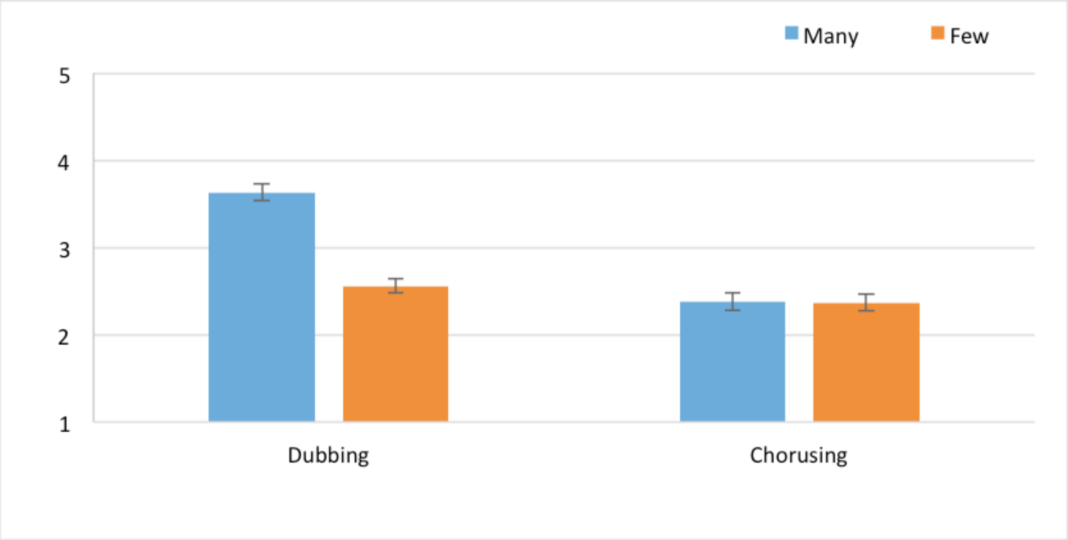


Figure 3.5. Ratings of amount stimuli. Error bars are standard error of the mean

The results are presented in Table 3.7 and illustrated in Figure 3.5. They are closely clustered around the mid-point of the Likert scale (3) with each one falling within a single standard deviation of it suggesting that neither technique was very effective. The mean Likert ratings for the stimuli that used the chorusing technique are almost indistinguishable from one another (.01) and the many stimuli was rated as having a below average number of components. This further suggests that the chorusing technique was not very effective. The results for the dubbing method were slightly better as the many stimulus was rated as having more components than the few stimulus by roughly a single Likert category (1.07).

Stimuli	**Mean Likert Rating**	**Std. Deviation**
Many Dubbing	3.63	1.12
Few Dubbing	2.56	0.95
Many Chorusing	2.38	1.16
Few Chorusing	2.37	1.18

Table 3.7. *Mean Likert Ratings and Standard Deviations for Amount Dimension*

The results were further analysed with a repeated measures ANOVA that was performed on the number measures with the design 2 (method: dubbing vs. chorusing) x 2 (number: many vs. few) with repeated measures on both factors. There was a main effect of method F(1, 138)=37.62, $p<.001$, $\eta_p^2 = .21$, such that listeners reported higher ratings for the dubbing method than for the chorusing method, the significant *p* value and large effect size suggest that this effect may be strong.

There was a main effect of number F(1, 138)=57.03, $p<.001$, $\eta_p^2 = .29$, such that participants rated the many stimuli as representing more elements than the few stimuli The significant *p* value and large effect size suggest that this effect may also be strong.

The method and number factors interacted F(1, 138)=41.78, $p<.001$, $\eta_p^2 = .23$. Contrasts to decompose the interaction indicated that listeners perceived more elements in the dubbing stimuli than the chorusing stimuli when the number of elements presented was few F(1, 138)=65.32, $p<.001$, $\eta_p^2 = .32$, but not when the number presented was many F(1, 138)=2.2, $p>.05$, $\eta_p^2 = .02$. This suggests a strong bias whereby listeners interpret stimuli created with the dubbing method as containing more elements than those created with the chorusing method when few elements are represented.

Listeners also perceived more elements in the many stimuli than the few stimuli for the dubbing method F(1, 138)=63.47, $p<.001$, $\eta_p^2 = .32$, but not for the chorusing method F<1. This suggests that the dubbing method is strongly effective, on the basis of *p* value and effect size, while the chorusing method is ineffective.

These results suggest that dubbing is a moderately effective method for influencing listeners' perception of the number of elements in a vocal gesture, but the chorusing method is ineffective. They also suggest that listeners have a bias whereby they rate stimuli which have the dubbing method applied as containing a larger number of elements than those with the chorusing method applied. Overall the results suggest that the dubbing method is of limited use in representing multiples of vocal gestures, while the chorusing method is ineffective.

3.24 Hypothesis (1) Evaluation D: Texture

This experiment is intended to determine if listeners perceive changes in the texture dimension of a synthesised vocal gesture and to establish the validity of two different methods for adding roughness to a vocal gesture. It was also intended to establish the validity of the synthesis parameters, derived through exploration with the prototyping platforms and compositional research, for mapping data to those dimensions. Johnson (1987) suggests that subjects develop a Rough-Smooth schema through frequent embodied encounters with processes and entities that change in texture in one's environment. As such, listeners should be able to identify changes in texture along a relative bi-polar scale from rough to smooth on the basis of this embodied schema. The relative synthesis parameters, found to be effective for modelling this dimension were determined using prototyping platforms and compositional research. The process by which these parameters were discovered is documented in appendix 3.4.

3.25 Design and Materials

Three stimuli were used in this experiment. One was expected to sound rough, another was expected to sound very rough and the last was expected to sound smooth. Exploration with prototyping platforms and research via composition determined the addition of noise to a smooth sounding stimulus as sufficient for controlling the perceived roughness of the stimulus. After taking part in a short training exercise to familiarise themselves with the evaluation procedure, listeners were asked to listen to each of the stimuli and to rate each one on a five point Likert scale of Very Smooth (1), Smooth (2), Neutral (3), Rough (4) and Very Rough (5).

The stimuli were created using additive synthesis techniques to synthesise vocal gestures. The first stimulus consists of a waveform with static amplitudes across 320 partials and an ‘O’ vowel profile with central frequency of 440Hz. This was expected to sound smooth. In the second stimulus, noise before, the amplitudes of the 320 partials were varied at random simulating the addition of noise to the original noiseless stimulus before the application of the vowel formant. This stimulus was expected to sound rough. In the third stimulus, noise after, a filter was applied after the vowel formant to create a beating effect that results in a more noisy sounding timbre. This stimulus was expected to sound very rough. Further details on the stimuli used are provided in appendix 3.5.

3.26 Results and Analysis

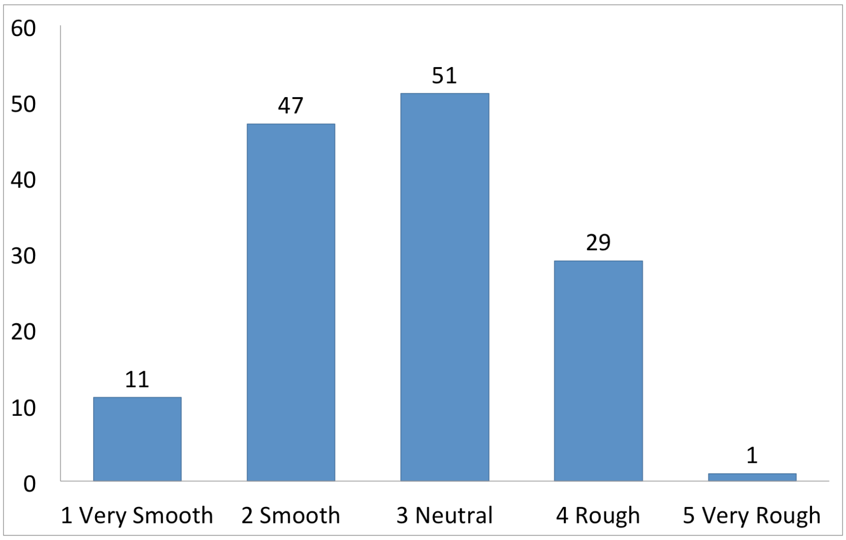


Figure 3.6. Listener Responses for Noiseless Stimulus

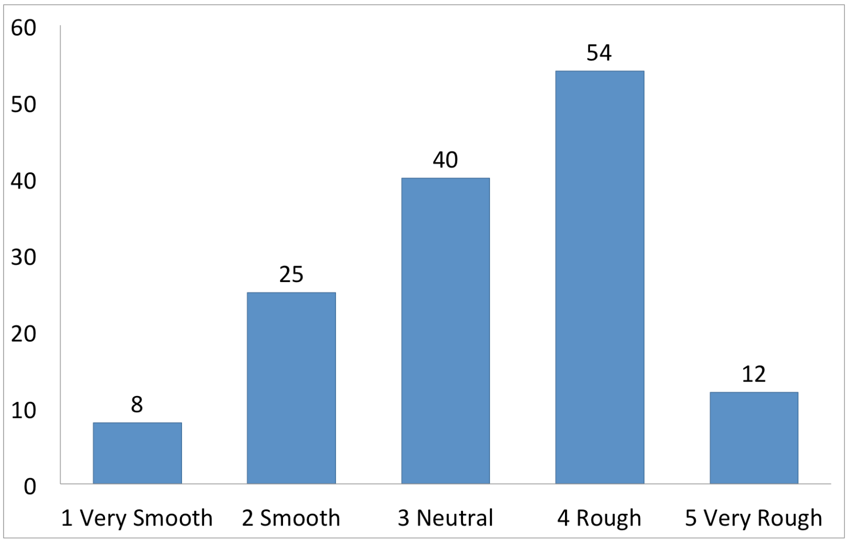


Figure 3.7. Listener Responses for Noise Before Stimulus

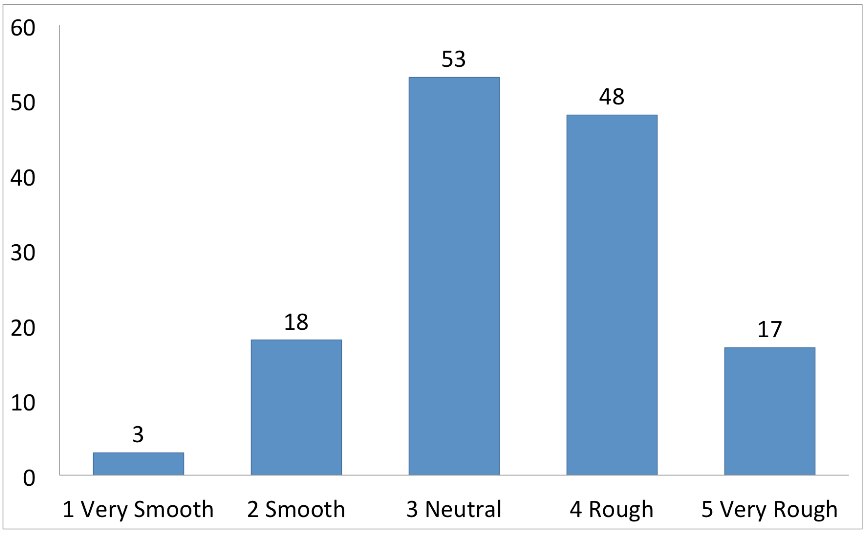


Figure 3.8. Listener Responses for Noise After Stimulus

The results are presented in Table 3.8. The results for the noiseless, noise before and noise after stimuli are illustrated in Figures 3.6, 3.7 and 3.8 respectively. The results clustered around the mid-point of the Likert scale (3) and each one falls within a single standard deviation of the mid-point. The mean Likert rating for the noiseless value placed it closer to rough (4) than smooth (2) and the mean Likert rating for the noise before stimulus placed it closer to smooth than rough. Figure 3.8 shows that 17 listeners rated the noise after stimulus as very rough and 48 rated it as rough but a small majority of listeners, 53, rated the stimulus as neutral. Figure 3.7 shows that 12 listeners rated the nose before stimulus as very rough and 54 rated it as rough but 40 rated the stimulus as neutral. Regardless, there is a much higher proportion of listener ratings clustering at the high end of the scale for noise after when compared to noise before. This suggests that the method of adding noise before the vowel formant might create a more smooth sounding timbre while adding noise after the vowel formant results in a more rough sound.

Stimuli	**Mean Likert Rating**	**Std. Dev.**
Noiseless	3.27	1.04
Noise Before	2.73	0.907
Noise After	3.42	0.939

Table 3.8. *Mean Likert Scale Ratings and Standard Deviations for Rough/Smooth Stimuli*

The results, illustrated in Figure 3.6 were analysed with a repeated measures ANOVA with the design 3 (texture: Noise Before vs. Noiseless vs. Noise After). There was a significant main effect of roughness F(2, 276)=22.38, $p<.001$, $\eta_p^2 = .14$, such that participants rated the stimuli in order of roughness as follows: noise after, noiseless, noise before. It was expected that the noiseless stimulus would be judged most smooth but the noise before stimulus was judged smoothest by the listeners. This may be because the addition of noise before the application of the vowel formant fills out the frequency spectrum allowing the formant profile to define a smoother sound. The significant *p* value and large effect size suggest that listeners perceive the noise after stimulus as substantially rougher than the noise before stimulus and the noiseless stimulus as substantially rougher than the noise before stimulus. In practical terms these results suggest that the method of adding noise before a vowel formant can be used to control the perception of smoothness and adding noise after the vowel formant can be used to control the perception of roughness in a vocal gesture but only to a limited extent.

3.27 Conclusions and Review of Hypothesis (1)

Hypothesis (1) proposed that listeners perceive embodied sonic dimensions of size, speed, amount and texture by applying the relevant embodied schematic organisation patterns to synthesised vocal gestures. Each of the previous evaluations aimed to determine whether or not listeners perceived the embodied dimensions in question and to determine how effective the synthesis parameters used were at controlling the listeners perception of these dimensions. The evaluations found that listeners reliably perceived each of these dimensions on the basis of the synthesis parameters used, but that the extent to which the synthesis parameters could control the listeners' perception of these dimensions was limited. Speed could be controlled to a moderate degree, while size, multitude and texture could be controlled to a small degree.

Although some of the results are robust with good levels of statistical significance, substantial effect sizes, and consistency across multiple evaluations, all of the Likert ratings recorded tended to cluster around the midpoint of the scale. This suggests that while the effects recorded were of practical significance, they may not present themselves as immediately obvious to the sonification listener in practice. Further research is required in order to determine the synthesis parameters that might be used to exaggerate these effects.

A number of sonifications were created in order to demonstrate the mapping of data to the dimensions isolated in the empirical evaluations for hypothesis (1). Speed and texture were mapped to the deprivation rate from 2007 to 2012 in two distinct sonifications. Size was mapped to the average unemployment rate per quarter from 2007 to 2012 and amount was mapped to the average emigration rate from 2007 to 2012. These sonifications were coded in the Csound audio programming language and the corresponding .csd implementations are included in the digital appendices for this chapter on the appendices DVD. The data was provided by the Irish Central Statistics Office.

3.28 Evaluating Hypothesis (2)

The following two evaluations were designed to test hypothesis (2) which holds that listeners perceive changes along dimensions of tension, weight, strength, texture, heat, size and brightness in vocal gestures and that these dimensions can be controlled using vowel formant profiles.

3.29 Hypothesis (2) Evaluation A: Attributes via Vowel Formant Profiles

Exploration with the prototyping platforms and research by composition suggested that listeners may relate vowel sounds to embodied schemata. This evaluation is intended to determine how strongly listeners relate seven unique vowel formant profiles, A, U, O, I, E, Ü, Ä, to the embodied attribute schemata Big-Small, Dark-Bright, Heavy-Light, Strong-Weak, Rough-Smooth, Hot-Cold listed by Macaranas *et al.* (2012). Johnson (1987) suggests that subjects develop a number of attribute schemata through frequent embodied encounters with changing physical attributes in one’s environment. As such, listeners should be able to identify changes in embodied sonic dimensions in reference to these embodied schemata. Exploration with the prototyping platforms suggested that listeners might associate different vowel formant profiles with different attribute schemata.

3.30 Design and Materials

Fourteen stimuli were used in this evaluation. Each of the stimuli featured a different vowel profile A, U, O, I, E, Ü, Ä. Two versions of each vowel sound, one with a distinct clearly pitched vocal timbre and one with a noisy vocal timbre, were created. A more in depth description of the synthesis parameters is available in the appendix for this chapter. Listeners were presented with each of the stimuli and asked to choose which pole of each of the six attribute schemata best describes that sound. Listeners were tasked with assigning a pole from each of the six schemata to each stimulus.

3.31 Results and Analysis

The results are presented in Table 3.9 and 3.10 below. Listeners categorised A as the strongest sounding vowel, and U as the weakest but while 81% of listeners categorised A as strong only 42% categorise U as weak. Listeners also categorised A to be the biggest sounding vowel and U to be the smallest but while 74% of listeners categorised A to be the biggest only 46% categorised U to be the smallest. Listeners categorised I as the brightest sounding vowel, and U as the darkest but while only 50% of listeners categorised I as bright 74% categorise U as dark. The results listed in Table 3.10 are clustered more closely around the 50% mark of the scale suggesting that listeners had difficulty relating the vowel sounds to attribute schemata. The average values show that 70% of the time listeners tend to interpret all vowel sounds as strong.

Vowel	**Strong**	**Weak**	**Big**	**Small**	**Bright**	**Dark**
A	81	19	74	26	39	61
U	58	42	54	46	26	74
O	60	40	63	37	39	61
I	73	27	60	40	50	50
E	73	27	69	31	49	51
Ü	65	35	59	41	32	68
Ä	78	22	68	32	45	55

Table 3.9. *Significant Vowel Attribute Results*

Vowel	**Heavy**	**Light**	**Rough**	**Smooth**	**Hot**	**Cold**
A	71	29	70	30	50	50
U	65	35	60	40	45	55
O	60	40	60	40	47	53
I	63	37	67	33	55	45
E	63	37	73	27	53	47
Ü	62	38	63	37	50	50
Ä	66	34	65	35	55	45

Table 3.10. *Non-significant Vowel Attribute Results*

The results were analysed by performing a repeated measures logistic regression on each of the attribute ratings (see Kleinbaum and Klein, 2010), with vowel (A ,U, O , I, E, Ü Ä) as the predictor variable, and listener categorisation (strong, weak, big, small, bright, dark, heavy, light, rough, smooth, hot, cold) as the respective dependent variables. This was intended to determine whether the number of listeners selecting negative and positive poles of each attribute differed between vowels.

The results for weight Wald F(6, 133)=.931, p>.05, Nagelkerke r^2 = .007, roughness Wald F(6, 133)=1.823, p>.05, Nagelkerke r^2 = .013, and heat Wald F(6, 133)=1.039, p>.05, Nagelkerke r^2 = .008, were non-significant suggesting that listeners do not associate these attributes with vowel sounds. The results for size Wald F(6, 133)=3.53, p<.05, Nagelkerke r^2 = .024 were strongly significant but only accounted for roughly 2% of the variance in listener response, indicating a small effect size. The results for brightness Wald F(6, 133)=3.95, p<.01, Nagelkerke r^2 = .039 were significant but only accounted for roughly 4% of the variance in listener response, indicating a small effect size. The results for strength Wald F(6, 133)=5.17, p<.001, Nagelkerke r^2 = .044 were strongly significant but only accounted for roughly 4% of the variance in listener response, indicating a small effect size.

The results suggest that vowel formant profiles can not be used to control the listeners perception of weak strength, small size, increased brightness, heavy and light weight, rough and smooth texture, or hot and cold temperature in a vocal gesture.
As such the results indicated vowel formant profiles cannot be used to control the listeners perception of strength along a scale from weak to strong, size along a scale from small to big, brightness along a scale from dark to bright, weight along a scale from heavy to light, texture along a scale from smooth to rough or temperature along a scale of cold to hot. These results do suggest that, to a limited extent, an A vowel formant profile can be used

to lend a sense of strength or a sense of large size to a vocal gesture and a U vowel formant profile can be used to lend a sense of darkness to a vocal gesture.

3.32 Hypothesis (2) Evaluation B: Tension via Vowel Formant Profiles

Exploration with the prototyping platforms and research by composition suggested that listeners may relate vowel sounds to tension. This evaluation is intended to determine how strongly listeners perceive dimensions of tension in vowel formant profiles, A, U, O, I, E, Ü, Ä. Johnson (1987) suggests that subjects develop a group of embodied schemata called force dynamic schemata from embodied experiences of physical forces. This group contains schemata describing compulsion, resistance and counter-force all of which are manifest in experiences of tension. As such, listeners should be able to identify changes in the embodied sonic dimension of tension in vocal gesture, in reference to these force dynamic schemata.

3.33 Design and Materials

Fourteen stimuli were used in this evaluation. Each of the stimuli features a different vowel profile A, U, O, I, E, Ü, Ä. Two versions of each vowel sound, one with a clearly pitched vocal timbre and one with a noisy vocal timbre, were created. The noisy timbres do not consist of pure noise with formant filters applied. They have a noisy sounding timbre but also have a central pitch. As such they can function as vowels. The pitched cues have a clear vocal timbre with a central pitch. A more in depth description of the synthesis parameters is available in the appendix for this chapter. Listeners were presented with all 14 stimuli and asked to rate each one on a 5 point scale of Very Relaxed (1), Relaxed (2), Neutral (3), Tense (4) and Very Tense (5).

3.34 Results and Analysis

Pitch Stimuli	Mean Likert Rating	Std. Dev.
A	3.69	0.962
U	3.11	0.858
O	3.28	0.814
I	3.82	0.95
E	3.84	0.945
Ü	3.16	0.88
Ä	3.54	0.885

Table 3.11. *Mean Likert Ratings and Standard Deviations for Pitched Tension Stimuli*

Noise Stimuli	Mean Likert Rating	Std. Dev.
A	2.97	0.852
U	2.92	0.959
O	2.91	0.962
I	3.35	0.984
E	3.12	0.927
Ü	3.1	0.961
Ä	3.08	0.974

Table 3.12. *Mean Likert Ratings and Standard Deviations for Noise Tension Stimuli*

The mean Likert ratings for the evaluation are presented in Table 3.11 and 3.12. All of the results recorded fall within one standard deviation of the midpoint of the Likert scale. This suggests that vowel profile may be of only limited effect in modelling tension. On the basis of decreasing tension listeners rated pitched vowels in the sequence E, I, A, Ä, O, Ü, U. The ratings for the pitched I and E stimuli are roughly the same being differentiated by only .02 of a Likert category. The ratings for pitched Ü and U stimuli are also roughly equivalent being differentiated by only .05 of a Likert category. On the basis of decreasing tension listeners rated noise based vowels in the sequence I, E, Ü, Ä, A, O, U. The noise based Ü and Ä stimuli were rated as roughly equivalent in tension being differentiated by only .02 of a Likert category. The noise based A and U stimuli were rated as roughly equivalent in tension being differentiated by only .05 of a Likert category, while U and O are only differentiated by .01 of a Likert category.

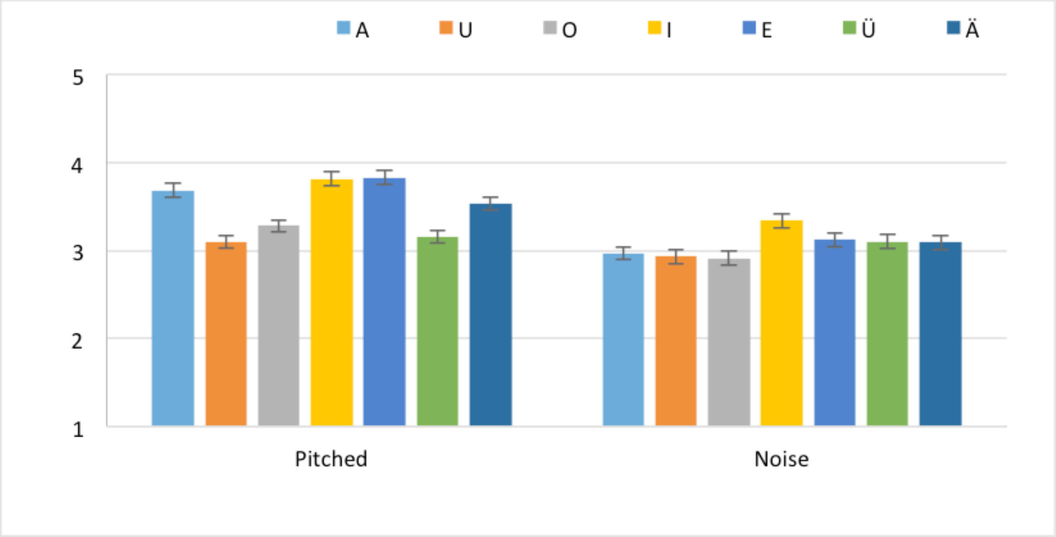


Figure 3.9. Listener ratings of Tension

The results, illustrated in Figure 3.9, were analysed using a repeated measures ANOVA with the design 2 (tone: pitched vs. noise) x 7 (vowel: A vs. U vs. O vs. I vs. E vs. Ü vs. Ä) with repeated measures on both factors.

There was a main effect of tone F(1, 135)=40.76, p<.001, $\eta_p^2 = .23$, such that pitched stimuli were judged more tense than noise stimuli. The highly significant p value and large effect size suggest that this is a practically significant result. There was a main effect of vowel F(6, 810)=19.71, p<.001, $\eta_p^2 = .13$, such that the vowels ranked from most to least tense were I, E, A, Ä, Ü, O, U. The highly significant p value and large effect size suggest that this ranking is reliable.

Tone interacted with vowel F(6, 810)=10.96, p<.001, $\eta_p^2 = .08$ and contrasts to decompose the interaction showed that there was a significant effect of vowel on both pitch F(6, 130)=19.32, p<.001, $\eta_p^2 = .47$ and noise F(6, 130)=4, p<.001, $\eta_p^2 = .16$. These results confirm that the sequences of descending tension for both pitch, E, I, A, Ä, O, Ü, U, and noise, I, E, Ü, Ä, A, O, U are of practical significance to sonification design, due to highly significant p values and large effect sizes. The larger effect size for pitched vowel stimuli also suggests that pitched vowel sounds are more effective than noisy vowel sounds for creating a sense of tension for a listener.

Pitched clips were perceived by listeners to be reliably more tense than noise clips for the vowels A, Ä, I and E respectively. The result for A was F(1, 135)=52.56, p<.001, $\eta_p^2 = .28$. The result for Ä was F(1, 135)=22.67, p<.001, $\eta_p^2 = .14$, the result for I was F(1, 135)=24.11, p<.001, $\eta_p^2 = .15$ and the result for E was F(1, 135)=58.35, p<.001, $\eta_p^2 = .3$. Less reliable were the results for U F(1, 135)=4.07, p<.05, $\eta_p^2 = .03$, which showed a small effect size and were just over the threshold of statistical significance. The pitched Ü stimulus was not reliably perceived to be more tense than its noise based counterpart F<1.

The results indicate that vowels formants profiles have a practically significant effect on perceived tension, particularly for pitched stimuli, however it must be noted that all of the stimuli were rated within one standard deviation of the mid-point of the Likert scale meaning that the effects, though reliably present, were not strong. The results suggest that the use of vowel formants to control the listener's perception tension in a vocal gesture is of limited effectiveness.

3.35 Conclusions and Review of Hypothesis (2)

Hypothesis (2) proposed that listeners perceive dimensions of tension, weight, strength, texture, heat, size and brightness on the basis of vowel formant profiles in synthesised vocal gestures. The first evaluation showed that listeners do not reliably perceive these dimensions on the basis of vowel formant profiles but that, to a limited extent, a sense of strength or largeness can be applied to a vocal gesture using an A formant profile and a sense of darkness can be applied using a U formant profile. The second evaluation of hypothesis (2) showed that, on the basis of vowel formant profiles, listeners could perceive tension in vocal gestures but only to a limited extent. As such further research might explore different ways of modelling the attribute schemata.

A number of sonifications were created in order to demonstrate how data can be mapped to the dimensions isolated in the empirical evaluations for hypothesis (2). Tension in a synthesised vocal gesture was mapped to the suicide rate from 2007 to 2012 wherein the lowest data value applied a U formant profile and the highest value applied an I profile to a vocal gesture. All values in between were mapped to a corresponding blend between a U formant and I formant profile. Size in a synthesised vocal gesture was mapped to the GNP rate from 2007 to 2012 wherein the lowest data value applied a U formant profile and highest value applied an A formant profile to a vocal gesture. For both sonifications, all data values in between the extrema applied mixtures of both formant profiles in question to a vocal gesture. These sonifications were coded in the Csound audio programming language and the corresponding .csd implementations are included in the digital appendices for this chapter on the accompanying appendices DVD. The sonification data was provided by the Irish Central Statistics Office.

3.36 Evaluating Hypothesis (3)

The following evaluation was designed to test hypothesis (3) which holds that listeners perceive changes along the embodied sonic dimensions of happiness, tenderness, anger, fear and sadness in synthesised vocal gestures, and that these dimensions can be modelled by synthesising Juslin and Laukka's (2003) 'Cross-Modal Patterns of Acoustic Cues for Discrete Emotions' presented in Table 3.13. The parameters derived, during exploration with the prototyping platforms, for modelling the emotional dimensions of vocal gestures are described in detail in the appendix for this chapter.

Anger	Fast speech rate/tempo, high voice intensity/sound level, much voice intensity/sound level variability, much high-frequency energy, high F0/pitch level, much F0/pitch variability, rising F0/pitch contour, fast voice onsets/tone attacks, and microstructural irregularity.
Fear	Fast speech rate/tempo, low voice intensity/sound level (except in panic fear), much voice intensity/sound level variability, little high-frequency energy, high F0/pitch level, little F0/pitch variability, rising F0/pitch contour, and a lot of microstructural irregularity.
Happiness	Fast speech rate/tempo, medium–high voice intensity/sound level, medium high-frequency energy, high F0/pitch level, much F0/pitch variability, rising F0/pitch contour, fast voice onsets/tone attacks, and very little microstructural regularity.
Sadness	Slow speech rate/tempo, low voice intensity/sound level, little voice intensity/sound level variability, little high-frequency energy, low F0/pitch level, little F0/pitch variability, falling F0/pitch contour, slow voice onsets/tone attacks, and **microstructural irregularity.**
Tenderness	Slow speech rate/tempo, low voice intensity/sound level, little voice intensity/sound level variability, little high-frequency energy, low F0/pitch level, little F0/pitch variability, falling F0/pitch contours, slow voice onsets/tone attacks, and **microstructural regularity.**

Table 3.13. *Summary of Cross-Modal Patterns of Acoustic Cues for Discrete Emotions*

3.37 Design and Materials

Five stimuli in total were generated for this evaluation. Each stimulus was designed to afford the emotional dimensions described by Juslin and Laukka’s Cross-Modal Patterns of Acoustic Cues for Discrete Emotions and listed in Table 3.13. Listeners were presented with each of the five stimuli and asked to choose which emotion best describes it from the following options: Tenderness, Happiness, Anger, Sadness or Fear.

3.38 Results and Analysis

The results are presented in Table 3.14. None of the stimuli were correctly identified by a clear majority of listeners. 41% of listener identified the stimulus with the happy parameters as sounding happy. 40% of listeners correctly identified the stimulus with the anger parameters as sounding angry. 40% of listeners also misidentified the stimulus with the sad parameters as sounding fearful, suggesting that the parameters for sadness are better suited to modelling fear.

Emotion Parameters	**Happy**	**Sad**	**Fear**	**Anger**	**Tenderness**
Happy Parameters	41	14	11	19	14
Sad Parameters	11	14	40	28	7
Fear Parameters	12	25	29	14	19
Anger Parameters	19	11	18	40	12
Tenderness Parameters	10	32	29	22	8

Table 3.14. *Participant Responses for Emotion (% agreement)*

The results were further analysed with a repeated measures logistic regression, with Parameters (Happy, Sad, Fear, Anger, Tenderness) as the predictor variable, and listener categorisation of emotion (Happy, Sad, Fear, Anger, Tenderness) as the dependent variable. This test determined the effect of parameters on the listeners' categorisation of the stimuli. Although the results were highly significant Wald F(16, 123)=6.26, $p<.001$, Nagelkerke $r^2 = .175$ they only accounted for roughly 18% of the variance in listener response, indicating a small effect size. This suggests that while the effects described by the results are reliably present, they are not very strong.

Overall the parameters for happiness and anger are only effective to a very limited extent while the parameters for sadness are similarly effective at modelling fear. The parameters for fear, and tenderness are ineffective. The only difference between the cues for tenderness and sadness in Juslin and Laukka's (2003) Cross-Modal Patterns of Acoustic Cues for Discrete Emotions is microstructural irregularity[13]. The cues for sadness have a measure of microstructural irregularity while the cues for tenderness do not. It is possible that listeners cannot detect microstructural cues as well in synthesised

---

[13] tiny inconsistencies in frequency, intensity and duration.

vocal gestures and it is also possible that the synthesis parameters used to synthesise microstructural regularity are insufficient.

Regardless, the empirical results indicate that the synthesis parameters, and possibly the auditory cues, used to control sadness are better suited to controlling fear and those used for modelling tenderness are better suited to modelling sadness. Listeners showed a level of ambiguity in distinguishing between fear and sadness for the actual fear stimulus itself. The sadness stimulus was identified much more strongly with fear than the fear stimulus and tenderness stimulus was strongly identified as sad. The tenderness stimulus was strongly identified with fear, and most weakly identified with tenderness. No other stimulus was identified with tenderness. Two possible reasons are here suggested for this. Either the listener doesn't tend to recognise the emotion of tenderness very well or the synthesis parameters used to model tenderness were not very effective. Overall these results suggest the synthesis parameters investigated in this evaluation are severely limited in their ability to control the perception of emotion in a vocal gesture for large and non-specific international audiences of sonification listeners.

3.39 Conclusions and Review of Hypothesis (3)

These results show that listeners did not effectively perceive dimensions of happiness, fear and anger in vocal gestures synthesised on the basis of Juslin and Laukka's Cross-Modal Patterns of Acoustic Cues for Discrete Emotions (Juslin and Laukka, 2003). Although the results invalidate hypothesis (3), listeners did perceive dimensions of happiness and anger to a limited degree. They also perceived fear to a limited degree on the basis of the parameters for sadness. Future research might focus on attempting to refine the parameters used or on determining new parameters to exaggerate these effects. Future research should also consider that these parameters may not represent the necessary means for controlling emotion in a synthesized vocal gesture and new parameters may need to be taken into account. Determining an explanation as to why the parameters for sadness were better suited to controlling fear might also provide fruitful, as this result seems to suggest that Juslin and Laukka's cues, which were developed to describe actual vocalisations, do not transfer well to synthesised vocal gestures.

A sonification was created for the purpose of illustrating how data might be mapped to dimensions of this nature in a sonification context. The average deprivation rate per quarter from 2007-2012 was mapped to the anger dimension of a synthesised vocal gesture wherein the cues for anger intensify as the data increases. This sonification was coded in the Csound audio programming language and the corresponding .csd implementation is included in the digital appendices for this chapter on the appendices DVD. The sonification data was provided by the Irish Central Statistics Office.

3.40 Evaluating Hypothesis (4)

The following evaluation was designed to test hypothesis (4) which holds that listeners perceive embodied sonic dimensions of proximity, vertical orientation, horizontal orientation and approach-passage-retreat in synthesised vocal gestures and that these dimensions can be modelled and controlled through spatial synthesis techniques.
It also provides empirical support for the validity of the parameters, derived through exploration with the prototyping platforms and compositional research, for synthesising and controlling those dimensions.

Johnson (1987) suggests that subjects develop a number of spatial schemata through frequent embodied encounters with phenomena that move through space in one's environment. As such, listeners should be able to identify changes in embodied sonic dimensions in reference to these embodied spatial schemata. The synthesis parameters, determined during exploration with prototyping platforms, for modelling and controlling the spatial dimensions of vocal gestures are presented in Table 3.15.

3.41 Design and Materials

The evaluation of the dimensions offered by stereo spatial movement was straightforward. Pitch was treated as an indicator of height as the conceptual metaphorical relationship between pitch and height has been widely discussed across the embodied musicological literature (e.g., Johnson, 2008; Brower, 2000; Zbikowski, 2005). Panning was use to give a sense of horizontal dimensionality and the Doppler effect was applied to give the listener the sense that a vocal gesture was approaching, then passing by and then retreating from a listener. The evaluation was simply intended to reaffirm the already well established finding that these spatial dimensions (Baier *et al.*, 2007) can be recruited in sonification. As such listeners were simply asked to identify the specific movements and locations of vocal gestures in a stereo spatial field. There is however a novel component in this evaluation, the application of the Doppler effect to a vocal gesture to recruit the dimension in which vocal gestures approach, pass by and retreat from a listener.

Fifteen stimuli in total were generated for this evaluation. These stimuli are represented in the Table 3.15. For stimuli where vocal gestures moved through stereo space, listeners were asked to identify the movement trajectories of the sounds from a drop down list. The stationary stimuli held a fixed location for their duration. Listeners were asked to rate these stimuli on a five point Likert scale in terms of vertical orientation, horizontal orientation or proximity. The simple movement stimuli changed spatial location across a single axis. For these stimuli, listeners were asked to select the correct movement trajectory from five possible alternatives. The complex movement stimuli changed spatial location across two axes simultaneously. For these stimuli, listeners were asked to select the correct movement trajectory from four possible alternatives.

Synthesis Dimensions	**Spatial Dimensions**
Doppler Effect	Approaching-Passing-Retreating
Panning	Left to Right
Panning	Right to Left
Vertical Pitch	High to Low
Vertical Pitch	Low to High
Pitch and Panning	High Right to Low Left
Pitch and Panning	High Left to Low Right
Pitch and Panning	Low Right to High Left
Pitch and Panning	Low Left to High Right
High Pitch	High
Low Pitch	Low
Panning	Right
Panning	Left
Reverb	Near
Reverb	Far

Table 3.15. *Stereo Spatial Parameters*

3.42 Results and Analysis

The results for the simple movement evaluation and the complex movement evaluation are presented in Tables 3.16 and 3.17 and the correct answers are highlighted in bold. The majority of listeners correctly identified the movement trajectories of each of the stimuli, with the exception of low right to high left. These majorities are still much lower than anticipated for both simple and complex movement trajectories. A small majority of listeners misidentified the low left to high right stimulus as moving from high left to low right, suggesting a measure of confusion on the part of the listeners for this stimulus.

Stimuli	**Left to right**	**Right to left**	**Up to down**	**Down to up**	**Front to back**
Left to right	**55**	18	16	8	3
Right to left	30	**39**	21	7	3
Up to down	20	9	**55**	10	6
Down to up	17	4	18	**52**	9

Table 3.16. *Simple Stereo Spatial Parameters Results*

Stimuli	**High left to low right**	**High right to low left**	**Low right to high left**	**Low left to high right**
High left to low right	**46**	25	16	13
High right to low left	30	**43**	15	12
Low right to high left	27	13	**44**	16
Low left to high right	36	13	20	**31**

Table 3.17. *Complex Stereo Spatial Parameters*

A repeated measures logistic regression, with simple stereo spatial parameters (left to right, right to left, up to down, down to up) as the predictor variable, and listener categorisation of simple trajectory (left to right, right to left, up to down, down to up, front to back) as the dependent variable was carried out. This analysis was performed in order to analyse the results for the simple movement trajectories. These results were found to be of high statistical significance Wald F(12, 127)=10.61, $p<.001$, Nagelkerke $r^2 = .353$ but they only accounted for 35% of the variance in listener response. In the context of the perception of simple stereo spatial movements, this represents a small effect size, as it would be expected that listeners would be more effective at determining spatial cues. This suggests that a substantial portion of the listeners cannot correctly identify the movement of the stimulus.

A second repeated measures logistic regression performed, with the complex stereo spatial parameters (high left to low right, high right to low left, low right to high left, low left to high right) as the predictor variable, and listener categorisation of complex trajectory (high left to low right, high right to low left, low right to high left, low left to high right) as the dependent variable was carried out. This analysis was performed to analyse the results for the complex movement trajectories. They were found to be highly significant Wald F(9, 130)=6.09, *p*<.001, Nagelkerke r^2=.156, but they only accounted for 16% of the variance in listener response. This effect size is roughly half the size of that recorded for the simple movement trajectories suggesting a very small effect size. This suggests that the majority of listeners cannot identify complex spatial movements applied to a vocal gesture.

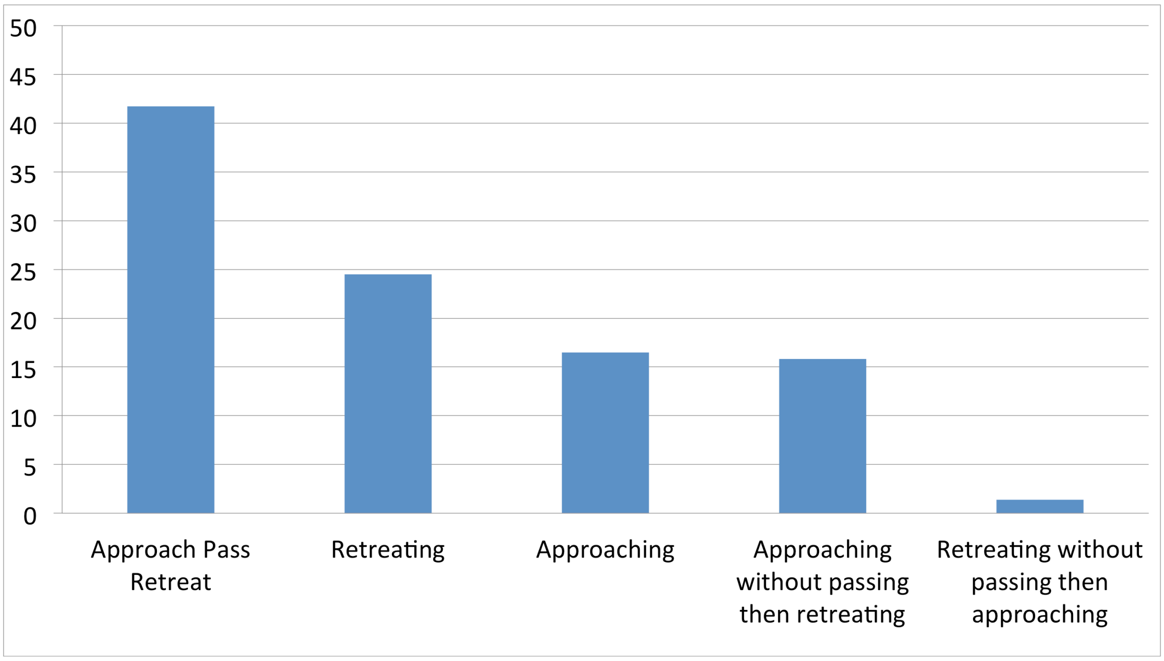


Figure 3.10. Results for approach-passage-retreat

The results for the approaching-passing-retreating stimulus are represented in Figure 3.10. 42% of the listeners accurately identified the movement trajectory. In similar fashion to the other results obtained for movement trajectories, this percentage is lower than anticipated. These results suggest that the majority of listeners could not identify the trajectory of a vocal gesture that appeared to approach, pass by and retreat from a listener on the basis of Doppler shift cues.

Stimuli	**Mean**	**Std. Deviation**
High	3.38	0.88
Low	2.48	0.854
Right	3.24	0.937
Left	2.58	1.014
Near	2.88	0.737
Far	3	0.971

Table 3.18. *Stationary Stimuli Results*

The results for the stationary stimuli are presented in Table 3.18. Listeners' ratings were recorded on five point Likert scales. For vertical orientation stimuli this scale ran from Very Low (1) to Very High (5). For horizontal orientation stimuli this scale ran from Far Left (1) to Far Right (5). For distance stimuli this scale ran from Very Near (1) to Very Far (5). All of the stimuli fell within one standard deviation of the mean, suggesting that effects were small. Listeners perceived the high stimulus as being higher than the low stimulus and vice versa and listeners also perceived the left stimulus as being positioned further to the left than the right stimulus and vice versa. Listeners interpreted the near and far stimuli as being positioned at roughly the same distance. This suggests that the parameters applied to the near and far stimuli were not effective at controlling a listener's perception of distance in a vocal gesture.

These results were further analysed using three paired sample t-tests which confirmed that listeners rated the high stimuli significantly higher than the low stimuli t(138)=9.22, *p*<.001, d= .78) and *vice versa*. The significant *p* value and large effect size, indicated by d, suggest a reliable effect of large practical significance. Listeners rated the left stimuli as significantly more to the left than the right stimuli t(138)=4.95*, p*<.001, d= .42) and *vice versa*. The significant *p* value and small effect size suggest a moderately reliable effect that is of small practical significance. Listeners did not distinguish between near and far stimuli with a high level of statistical significance t(138)=1.39, *p*>.05. This may be because no context was provided for a listener to distinguish a near vocal gesture from a far vocal gesture and *vice versa*.

3.43 Conclusions and Review of Hypothesis (4)

Hypothesis (4) proposed that listeners perceive dimensions of proximity, vertical and horizontal orientation and approach-passage-retreat in a vocal gesture on the basis of synthesised spatial cues. On the basis of the spatial cues used, listeners perceived vertical and horizontal orientation and approach-passage-retreat in a limited manner but did not reliably perceive proximity. These weaker than expected results might be a result of the vocal gesture stimuli to which the parameters were applied. Stimuli with richer spectral information may have performed better in the proximity and approach-passage-retreat tests but should not have performed much better in the vertical and horizontal orientation tests. The vocal gesture stimuli had vowel profile formants applied which slowly changed shape over the duration of the stimuli and this change in vowel formant shape may have interacted with the listeners' perception of spatial cues causing confusion. It is also possible that the choice of five possible answers was confusing to the listener. Furthermore the proximity cues might require additional context by which a listener can perceive the relative distance of a stimulus.

Further research might focus on refining spatialisation parameters that are effective for the dynamically evolving formant profiles inherent to vocal gestures and also explore alternative experimental designs. A sonification was created for the purpose of illustrating how data might be mapped to these spatial parameters to sonify the number of controlled drug offences from 2007-2012, using the Csound audio programming language is included in the digital and written appendices for this chapter.

3.44 Discussion

This chapter examined the embodied sonic dimensions offered by the embodied sonic complexes of vocal gestures. It aimed to empirically verify whether the hypothetical dimensions uncovered during exploratory research with the prototyping platforms and compositional process could be used to effectively communicate with a listener. It also aimed to determine approaches that could be used to effectively synthesise these dimensions.

The results generated in these evaluations are preliminary in nature and many of the effects recorded translate to small perceptual results in terms of sonification listening. The results for the hypothesis (1) evaluations suggest that the approaches to controlling the perceived size, speed amount and texture of a vocal gesture explored in evaluation 1 show promise and are worthy of further investigation.

The results for the hypothesis (2) evaluations suggest that vowel formant profiles cannot be used to effectively control a listeners perception of weight, strength, texture, heat, size and brightness in vocal gestures. They can be used however to add a sense of strength, large size, and darkness to a vocal gesture and to control the perceived sense of tension in a limited manner. The limited extent of these effects suggest that vowel formant profiles alone may not provide the necessary means for controlling these embodied sonic dimensions and that more parameters may need to be controlled in order to effectively achieve this. For example, when attempting to control size, the parameters determined in evaluation 1 could be used alongside vowel formants to improve the listener's perception of changes in the size dimension.

The results for the evaluation of hypothesis (3) also invalidated the hypothesis that Juslin and Laukka’s Cross-Modal Patterns of Acoustic Cues for Discrete Emotions provide the necessary means by which to control the listeners' perception of emotion in a vocal gesture showing instead that anger and happiness work to a very limited extent, the cues for sadness are better suited to controlling fear, and the cues for fear and tenderness are ineffective. These results suggest that the perception of emotion in a synthesised vocal gesture is a complex operation and reliant on more detailed parameters than those explored in this evaluation.

Finally, the results for the fourth evaluation showed that a listener's perception of the vertical and horizontal orientation and the approach-passage and retreat of a vocal gesture can only be controlled to limited extent while proximity cannot be effectively controlled at all, on the basis of the synthesis parameters used.

Although these results are preliminary in nature they do suggest embodied sonic dimensions can be controlled for the purposes of data sonification and they have also helped to determine the avenues of future research that need to be explored in order to make the use of embodied sonic dimensions a reality for sonification. These embodied sonic dimensions can be used to sonify data recorded along similar measurement dimensions. For example data relating to size, weight, speed *etc.* can be expressed using the size, weight or speed dimensions provided by synthesised vocal gestures and described in this chapter. This would make for intuitive, easy to understand sonifications that do not require extensive training periods on behalf of the listener before interpreting a sonification but rather exploit the embodied meaning-making to effectively communicate data to the listener.

According to the ESLM, during sonification listening embodied sonic dimensions become metaphors for the measurement dimensions of the original data. Sonifying data measurements by mapping them to analogous embodied sonic dimensions is consistent with this model. In the case of socio-economic data, the data source is the specific group of people measured. The data dimension is the socio-economic measurement (e.g., unemployment, emigration) taken of that group. The intent of *The Human Cost* was to communicate data about people. The main channel by which people communicate with one another is the voice. As such the voice provided an appropriate and direct sonic metaphor for the original data source: the people measured.

A number of strategies by which approaches to sonification informed by embodied cognitive science can be applied to aesthetic practices in the context of data-driven are represented in *The Human Cost*. This piece was chosen for performance at the Kevin Barry Room of the Irish National Concert Hall during the Contemporary Music Centre's

2015 Spring Salon. This was the first salon session by the CMC to focus solely on electronic music and the first performance at a CMC salon to include sonification techniques.

Based on the communicative effectiveness of the embodied sonic dimensions uncovered in this chapter, it is suggested that the human voice in general, and vocal gestures in particular, provide an appropriate domain of sonic complexes for both communicatively rich and effective sonification as well as aesthetically engaging data-driven composition. It is also suggested that mapping data dimensions to incompatible embodied sonic dimensions might be favourable when those dimensions can communicate extra layers of meaning that enrich the sonification. Whether or not this is actually the case must be decided on a case-by-case basis through expert opinion and user testing for sonification practice, or the appraisal of artistic and aesthetic merit for data-driven composition. The embodied sonic dimensions demonstrated in this chapter may potentially provide an elegant solution to both aspects of the mapping problem.

The first aspect of the mapping problem discussed in this chapter asks how best to map data to sound. This can be addressed in a manner that embraces embodied meaning-making by mapping data to the sonic complexes and dimensions afforded to a listener by their embodied nature. The second aspect asks how the perceptual entanglement of non-linear acoustic dimensions can be dealt with in sonification. Embodied sonic dimensions deal with this problem by treating only the cohesive and continuous aspects of embodied sonic experience as dimensions. These are significant contributions, as the mapping problem represents one of the biggest obstacles to the development of sonification research and practice.

A number of embodied sonic dimensions for the sonic complexes of vocal gestures are presented in this chapter. Researchers and practitioners can find new embodied sonic dimensions by applying the research techniques laid out in this chapter.

Researchers and practitioners should begin with a thorough reading of the relevant embodied cognition and sonification literature. Following this, candidate dimensions can be determined through exploration of hypothetical dimensions developed in the literature review. This can be achieved with prototyping platforms and the use of data-driven composition as a research method. These candidate dimensions should then be submitted to listener testing to determine their communicative effectiveness.

3.45 Summary

This chapter examined the embodied sonic dimensions, as discussed in the ESLM, of vocal gestures. It introduced and defined the vocal gesture before discussing some of the research techniques, prototyping platforms and research via composition, used to explore the dimensions that vocal gestures offer for sonification. It isolates, defines and empirically evaluates the communicative effectiveness of a number of these dimensions within the domain of embodied sonic complexes referred to in this chapter as vocal gestures. It introduces and explains the concepts of vocal gestures and stereo space before discussing two prototyping platforms developed to explore the kinds of embodied sonic dimensions offered by vocal gestures. A piece of data-driven music entitled *The Human Cost* is then introduced and discussed. The compositional process for this piece, alongside exploration with the prototyping platforms, was used as a method for further exploring the embodied sonic dimensions offered by the domain of vocal gestures. A number of empirical listener evaluations intended to test the communicative effectiveness of these dimensions are then presented alongside sonifications created using the dimensions. This is followed by a discussion of the work presented in the chapter and its implications for sonification research and practice.

Chapter 4: Embodied Sonic Complexes – Embodied Soundscape Sonification Framework

4.1 Introduction

Chapter 4 presents the Embodied Soundscape Sonification Framework, which adopts environmental soundscapes as a domain of embodied sonic complexes for sonification and arranges those sonic complexes in order to exploit the embodied nature of auditory cognition for communicatively effective sonification. The chapter opens with a consideration of embodied soundscape sonification, examining the embodied cognitive underpinnings of soundscape listening and discussing the relationship between soundscape and sonification. A set of prototyping platforms used to investigate the new dimensions and possibilities offered by soundscape sonification is presented and discussed. This is followed by the presentation of a number of soundscape sonifications of socio-economic data that were designed around principles derived through experimentation with the prototyping platforms. During the exploratory and compositional stages a number of embodied soundscape sonification techniques are uncovered. These techniques are combined to create four candidate frameworks for embodied soundscape sonification. Each framework is submitted to an experimental listener evaluation intended to test its communicative effectiveness for embodied soundscape sonification. The results of these evaluations are presented and analysed and a successful framework is chosen. The chapter closes with a discussion of the findings from this experimental investigation.

4.2 Embodied Soundscape Sonification

Environmental soundscapes, the environmental variant of the soundscapes defined by Schafer (1969), explored in the work of Truax (1996; 2001; 2012) and Westerkamp (1999; 2007) and extended to sonification by Walker and Kramer (2004), provide a communicatively effective domain of sonic complexes for sonification. This chapter demonstrates that soundscape sonifications can be organised and presented to a listener in ways that exploit their embodied cognitive meaning-making faculties.

Schafer (1969) originally defined the soundscape as any perceptible ‘sonic environment’, geographical, musical or otherwise which is distinguished from the wider acoustic environment which contains further non-perceptible acoustic components. The Handbook of Acoustic Ecology (Truax, 1978) further defined the soundscape as follows:

> An environment of sound (sonic environment) with emphasis on the way it is perceived and understood by the individual, or by a society. It thus depends on the relationship between the individual and any such environment. The term may refer to actual environments, or to abstract constructions such as musical compositions and tape montages, particularly when considered as an artificial environment.

Schafer's definition of the soundscape as the experience of the perceptible signals in an acoustic environment carries the subtle but important distinction that the soundscape results from the filtering and interpretation of the acoustic environment by the perceptual apparatus of the perceiving agent. The second definition presented here carries this further by explicitly highlighting the role of the listeners understanding as crucial to the definition of a soundscape. Truax (1996; 2012) argues that the relationships between human organisms and their environments are primarily mediated by soundscapes through dynamic processes of embodied cognition. A similar argument about the relationship between environment and organism is put forward by Maturana and Varela (1987) and Varela *et al.* (1991). They argue that the organism and environment share a close relationship of mutual specification or structural coupling. This is a closed loop relationship in which the environment defines and shapes the organism's perceptual and cognitive functions, through the conduit of the physical body, and those functions also define the organism's experience of the environment at the levels of perception, cognition and interaction, through the conduit of the physical body, in equal measure. This differs from Gibson's ecological view of perception in its focus on the physical body as the conduit between mind and world. The environments ability to shape the cognitive faculties of the organism is constrained by and mediated in terms of the physical body and the organism's ability to shape the environment is constrained by and mediated in terms of the physical body. Sound localisation by interaural time and interaural level difference[14] provides a relevant example of how the physical body's constraint and mediation of the environment defines the organism's cognitive faculties in terms of the listener's ability to determine the location of a sounds source.

Soundscapes are embodied because they are the experience that results from the constraint and mediation of the acoustic environment by the listener's physical body. This

---

[14] Interaural time and level differences refer to the differences in arrival times and arrival amplitudes of a sound between two ears.

suggests that the embodied cognitive capacities discussed in chapter 1 might be at play in the listener's perceptual organisation and conceptual understanding of the soundscape. Kendall (2010; 2014) has argued that the embodied cognitive faculties of embodied schemata and conceptual blending are crucial to the emergence of meaning in soundscape listening in the context of electroacoustic music. Droumeva *et. al.* (2009) have demonstrated how the experience of a soundscape and its conceptual interpretation can be structured on the basis of conceptual metaphors and embodied schemata in an auditory display context. For a sound to be embodied it need not represent some aspect of the human body and it need not sound as though it originated with the human body either. This thesis argues that auditory cognition itself is embodied and as a result all sounds are understood and made meaningful in terms of a mind that has been defined and shaped by the physical human body. The embodied approach advocated here involves using and controlling sounds in ways that exploit those aspects of cognition, which have been shaped by the body; it is the nature of cognition, and not the sounds themselves, that is considered embodied. This chapter examines organisational strategies by which soundscapes can be configured to best exploit the embodied meaning making capacities of the listener in a sonification context.

The term soundscape sonification is used in this thesis to refer to the use of recordings of real-world locations, the use of soundscapes which have been composed from recording of real world environmental elements, and to the use of soundscapes which have been composed from synthesised elements in the sonification of data. A number of sonification researchers and practitioners have explored and applied similar concepts of soundscape sonification (Gaver, 1993; Fernström and Brazil, 2001; Kilander and Lönnqvist, 2002; Hermann and Ritter, 2004, Mauney and Walker, 2004; Walker and Kramer, 2005; Walker *et al.*, 2006; Brazil and Fernström, 2007; Nees and Walker, 2009; Polli, 2012; Wolf, Gliner and Fiebrink, 2015; Hermann, *et al.*, 2015). Walker and Kramer (2004) recognise the importance of environmental sound to sonification. They argue that traditional psychoacoustics does not offer a comprehensive description of how listeners hear and think about real world sound and suggest that an ecological psychoacoustics that considers the relationship between organism and environment is required for the design of effective sonifications. This chapter further argues that a deeper understanding of the role of the body as the mediating and constraining conduit between the organism's cognitive faculties and environment is required to better understand meaning making in the context of soundscape sonification.

An embodied soundscape sonification is here defined as a sonification in which environmental soundscape elements are organised, mapped to data and presented in a manner that exploits the embodied nature of auditory cognition as described by the Embodied Sonification Listening Model (ESLM). The work presented in the remainder of this chapter assembles a design framework for creating embodied soundscape sonifications.

4.3 Prototyping Platforms and Exploration

Six prototyping platforms were developed using the Csound programming language to explore different approaches to synthesising embodied soundscape sonification techniques for creating effective data-driven soundscapes. One group of prototyping platforms were designed to be literal analogues of the data they would represent. They were simulations of different sound sources that could be mapped to data and used as sonic elements in a soundscape. A wind and a rain model were developed for expressing rainfall and wind speed data. A fire model, which is arguably less of a literal analogue, was developed for expressing temperature data.

A second group was metaphorically related to the data they were intended to sonify on the basis of Lakoff and Johnson's (Lakoff and Johnson, 1980) conceptual metaphor theory. A flowing water model was developed to represent emigration. This was informed by a number of studies which suggest a conceptual metaphorical mapping where the concept of a flood is used to reason and make sense of emigration (Dervinytė, 2009; Santa Ana, 1999). A clock model was developed to express financial data. This was informed by Lakoff and Johnson (1980; 1999) who suggest that people reason about money in terms of time as highlighted in the 'Time is Money' idiom. A heartbeat model was developed to represent economic data. This was informed by accounts which suggest that a conceptual metaphorical mapping exists whereby people think about the economy in terms of a living organism and reason about the number of goods and services produced in an economy in terms of biological growth and decay (White, 2003; Hodgson, 1993). The prototyping platforms do not represent direct physical models but rather attempts at simulating the sounds in question. These prototyping platforms are presented in the digital appendices for this chapter on the appendices DVD. Exploration with the prototyping platforms highlighted a number of shortcomings.

The sounds created by the prototyping platforms did not sound convincingly real, making it possible to create only rudimentary soundscapes using the platforms. Mapping data sets to individual prototyping platforms and arranging their outputs together in the style of a soundscape made for incoherent results. It became increasingly difficult to follow what was happening in the individual data sets as the number of data sets increased. It was decided that the soundscapes created using the platforms did not sound like they were emanating from a shared acoustic environment. They resembled chaotic arrangements of environmental sounds. This may have been caused by the emergence of conflicting or unrealistic psychoacoustic cues when the sonic outputs of each platform were grouped together. The mapping of each platform to a unique data set contributed to the chaotic unintelligible nature of the sonifications. These factors could be attenuated to a limited degree by mapping all of the prototyping platforms in a specific sonification to a common data set and by distributing their sonic outputs across the stereo spatial field. The application of formal structure derived from Johnson's (1987) near-far, left-right, centre-periphery, balance and twin-pan balance embodied schemata, to the soundscapes created boosted the listener's comprehension of the data expressed during informal testing. Although the sounds synthesised were not convincingly real they could be made more aesthetically interesting by mixing them to create new timbres. One such mixture between a vocal like sound and a siren like sound was used as the basis of the *Idle Hands* composition discussed later in this chapter. Exploration with the prototyping platforms suggested that soundscape could provide an effective avenue for sonification.

4.4 Data-Driven Soundscape Composition

During exploration with the prototyping platforms the aesthetic potential inherent in the mixture of ecological sonic sources with embodied sonic dimensions became apparent. *Idle Hands: A 31-part exploration of Irish Unemployment from 1983-2014 in G major* (hereafter *Idle Hands*) is a soundscape sonification that expresses Irish Central Statistics Office's Standardised Unemployment Rate for period from 1983 to 2014. A number of sonification researchers and arts practitioners have employed sonification techniques to create data-driven music compositions that raise awareness about social and environmental issues (Barrett and Mair, 2014; Polli, 2012; Taylor and Fernström, 2012; Quinn, 2001). *Idle Hands* continues this in this tradition. The piece represents an early step along the road to create a realised embodied soundscape sonification framework.

The majority of the piece was created in Csound. It uses granular synthesis techniques applied to a WAV soundfile which was created using fof synthesis techniques to produce a series of low pitched impulses at a low grain rate. The piece consists of data-driven analogues of three elements introduced by Schafer (1969) and discussed across the soundscape literature. These are the sound signal the soundmark and the keynote sound. These are discussed in greater detail in appendix 4.2. The sound signal is heard shifting between the middle ground and foreground as an imaginary hybrid of a distressed human vocal-like timbre and an alarming siren-like timbre that increases and decreases in pitch in response to the data. This sound signal is mapped to pitch so that the pitch increases as data values increase and decreases as they decrease. This timbre was selected to reflect the social and human cost associated with high unemployment levels. The soundmark is heard shifting between the foreground and middle ground as a timbral texture of evolving patterns. The soundmark is determined by the density of the grain cloud, which is driven by the data. The keynote is provided by the application of a harmoniser, a bank of tuned filters intended to extract a specific chordal configuration from an audio signal, to the sound signal and soundmark. The keynote presents a static benchmark in the piece against which the changes in the sound signal and soundmark can be compared. Reverb modelling is manually applied at a number of points in the composition process to help exaggerate timbres and further shape the sonic materials.

Idle Hands relies on conceptual metaphor (Lakoff and Johnson, 1980) to link sonic complexes to a data source and is structured around Johnson's Source-Path-Goal schema with the signal representing a trajector that navigates from a temporal source point of 1983 along its path, the intervening years, to reach its final destination in 2014. *Idle Hands* was chosen for performance at the Irish Sound Science and Technology Convocation at The University of Maynooth in August 2014.

Doom & Gloom is a soundscape sonification that further explores the human cost of Ireland's economic crash. The piece harnesses soundscape sonification techniques to present a study of Irish stereotypes in the pre-crash and post-crash eras. Historically, sonification has been used for the expression of quantitative data with too much of the focus in the area being on the development of supposedly objective dimensional specifications for expressing data. This work rejects the idea of objective frameworks for expressing qualitative data and proposes that it is precisely sound's ability to capture and communicate highly specific meanings about concrete qualitative phenomena, which make it a powerful medium for the expression of data.

Doom & Gloom expresses Irish Central Statistics Office's GNP (Gross National Product) and Standardised Unemployment rates from 2007 to 2012 and represents an early step along the path to creating a metaphorical twin-pan balance soundscape sonification as described later in this chapter. Two soundscapes are used in the creation of this piece. The first is a soundscape derived from Francis Ford Coppola's *Finian's Rainbow* where a Leprechaun challenges the thief who has gone to America with his pot of gold. Without the gold Ireland has fallen to rack and ruin. The montage presents a fitting metaphor for the Irish stereotypes prevalent in the pre-crash climate. This soundscape offers a heavily sardonic parody of the bleak and desolate future looming on the horizon for an Ireland that would lose its own metaphorical pot of gold. In short order this comical stereotype dissolves into a chaotic swarm-like grain cloud of contorted voices. Fog synthesis techniques (an extension of fof synthesis to soundfile granulation) granulate the original soundfile while the frequency of grains per second and transposition factor of the resulting grain cloud as well as the duration of each individual grain is driven by the Central Statistics Office's measured decline in the Irish GNP (Gross National Product) during the crash.

A second soundscape is introduced into the grain cloud already in its granulated form via fog synthesis techniques. This rematerialises and resolves into a stable soundscape at a rate defined by the unemployment data as applied to the grain rate, transposition rate and individual grain length of the fog algorithm. This soundscape is derived from an interview with an unemployed electrician named Christopher from an Al Jazeera documentary entitled *People & Power – Collapse of the Celtic Tiger* made by filmmaker Sinead O'Shea[15]. Christopher discusses his experiences of recent economic hardships and offers a first hand glimpse into the daily lives of the involuntarily unemployed, exposing some of the human factors that underlie government fact sheets and statistical data analyses on unemployment.

The entire piece is structured around Johnson's (1980) twin-pan balance schema with the balance slowly shifting as the sound shifts from left to right across the stereo space over the course of the piece. The piece uses this embodied schematic structure to present a contrast of attitudes to economic hardship in Ireland from the pre-crash to post-crash eras. The flippant notion of the happy go lucky Irish mischief-maker is audibly shattered to pieces by the decline in Irish GNP and rise of unemployment leads to a new, more resonant, kind of stereotype, the jobless and depressed twenty-something with no real prospects or hope for the future.

---

[15] published at aljazeera.com on the 19th of January 2012 http://www.aljazeera.com/programmes/peopleandpower/2012/01/201211984746688436.html

This piece is representative of what has been described as the “Aesthetic Turn” (Barrass, 2012) in sonification. This movement is a renaissance of sorts for the area, where composers are turning away from older failed paradigms and attitudes that sought to objectify and reduce the qualitative nature of sound. Instead they are extending sonification techniques and harnessing the full potential of sound to shape the cognitive processes of qualitative and aesthetic meaning-making. Much of this movement has been centred on the aesthetic sonification of environmental, scientific and socio-economic data as a new form of cultural medium.

Doom & Gloom has been chosen for performance at the following venues: The 2015 International Conference on Auditory Display ant Graz Austria, the Contemporary Music Centre’s 2015 Spring Salon at the Kevin Barry room of the national concert Hall in Dublin, the 2015 Sound and Music Computing Conference at Maynooth and the 2015 Irish Sound Science and Technology convocation at Dance Limerick.

A previous notable example of the use of sonification techniques in music composition is presented in David Spondike’s *Schnappschuss von der Erde* (Spondike, 2006). While *Idle Hands* focuses on a single socio-economic data set and *Doom and Gloom* uses two such data sets *Schnappschuss von der Erde* makes use of a large number of socio-economic data sets. The data set for the piece consists of 23 indicators from 187 countries between 2000 and 2004 and is provided by the World Bank. While *Schnappschuss von der Erde* uses the conventions of Western tonal system *Idle Hands* and *Doom and Gloom* are driven by the concept of the soundscape and make use of electroacoustic compositional techniques. However, both *Schnappschuss von der Erde* and *Doom and Gloom* use spatial transformations to represent data to a listener and all three pieces use changes in timbre to communicate data to the listener. Alongside spatialisation approaches *Schnappschuss von der Erde* uses melodic patterns and amplitude and timbral changes to create musical interest but *Idle Hands* and *Doom and Gloom* focus on evolving timbral patterns and modulating pitch as a means of creating musical interest. The compositional method used for *Idle Hands* and *Doom and Gloom* is very different to the method used for *Schnappschuss von der Erde.* This piece is divided into three sections each containing a number of individual sonifications. The sonifications were created using Symbolic Composer to map the data tables to control MIDI parameters. Order of presentation, pitch and loudness are the musical variables controlled by the data. This allowed for the creation of musical interest through to melodic patterns, amplitude and timbre changes and spatialisation within an eight-speaker array. A large number of sonifications were created but those included in the finished piece were chosen from a pool of fifty aesthetically pleasing sonifications. The compositional method used for *Idle Hands* and *Doom and Gloom* is described in the previous pages and further detailed in appendices 4.2 and 4.3. It differs greatly from the approach used in *Schnappschuss von der Erde.* Sonifications are created using Csound and then arranged further in Logic Pro X. Neither of these compositions make use of MIDI opting instead for fof and fog synthesis techniques. Furthermore, *Idle Hands* and *Doom & Gloom* were composed with embodied cognition principles in mind while *Schnappschuss von der Erde* was composed with the aesthetic potentials of the tonal system in mind.

4.5 Candidate Frameworks for Embodied Soundscape Sonification

Drawing from the exploration conducted with the prototyping platforms, and insights gained through the composition of *Idle Hands* and *Doom & Gloom*, four potential frameworks are here proposed and specified as candidates for an effective embodied soundscape sonification framework. These frameworks utilise two approaches to relating an ecological sonic source to a data set and two approaches to the formal organisation of those sonic sources to make them meaningful for the listener's embodied cognitive meaning-making faculties.

The first way of relating a sound source to a data source is here termed the metaphorical technique. It exploits the listener's faculty for conceptual metaphorical mapping. The second, here termed the blending technique, exploits the listener's faculty for conceptual blending.

In the metaphorical approach, illustrated in Figure 4.1, designers make use of ecological sonic complexes that are metaphorically related to the source of a data set. An example of the metaphorical method is the use of a soundscape of stormy weather to represent hard times because bad weather is often used as an embodied conceptual metaphor for hardship (Żołnowska, 2011). This approach differs from standard PMSon in that the data to sound mapping is mediated by an embodied conceptual metaphor, which is thought to work by the processes discussed by Lakoff and Johnson (1980). It is an embodied approach because conceptual metaphors are said to derive from previous embodied experience and are used to lend structure to novel experiences as discussed in detail in Chapter 1.

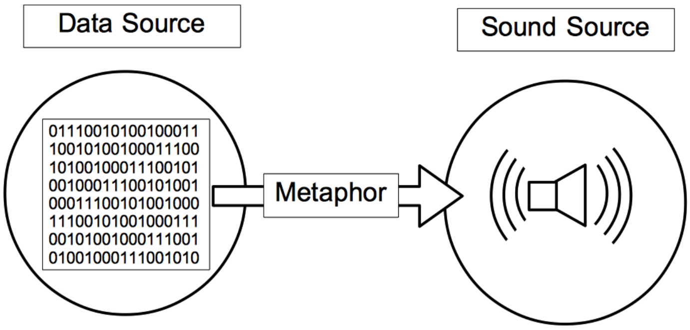


Figure 4.1. The Metaphorical Technique

In the blended approach, illustrated in Figure 4.2, designers create a soundscape that represents the imagined shared environment in which the data source is revealed by its causal effect on an independent sound source. An example of the blended method is the use of increasingly agitated birdsong to represent the increasing magnitude of an earthquake. In this blended soundscape the sound source, the birds, and the data source, the earthquake, co-exist in the same space. The magnitude of the earthquake, the data dimension, determines the magnitude of agitation of the birds, the sonic dimension. This differs from the metaphorical approach because it is the imagined causal relationship between the data-source and the sonic environment that the designer leverages to link data to sound. This approach differs from standard PMSon in that the data to sound mapping is designed to represent a conceptual blend, which is thought to work by the processes discussed by Fauconnier and Turner (2002). It is an embodied approach because conceptual blends are said to derive from previous embodied experience and are used to lend structure to acts of creative thought as discussed in detail in Chapter 1.

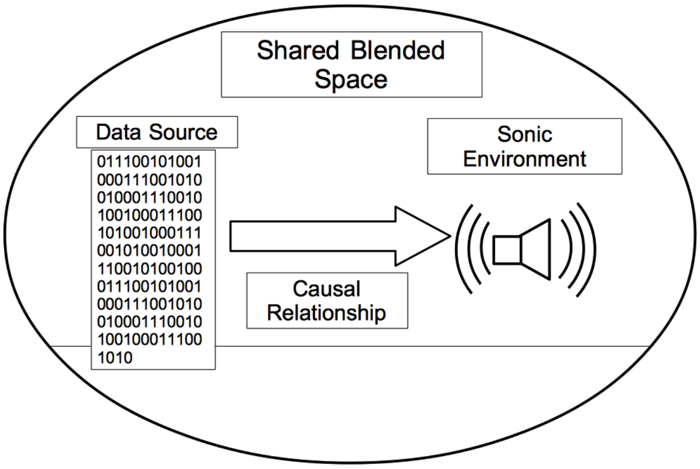


Figure 4.2. The Blended Technique

This chapter suggests that the formal embodied structure of a soundscape sonification can be arranged in at least one of two ways to help support the meaningful communication of a data set to a listener. There are also two methods for applying formal embodied structure to the sonification. These are the balance method and the twin-pan balance method. The first method of applying formal embodied structure is based on an embodied schema called the balance schema proposed by Johnson (1987). According to Johnson the balance schema is a formal pattern that humans derive from their first-hand experiences of learning to balance their bodies within their physical environments. A soundscape can be conceptualised as a set of unique individual sonic elements that bear a relationship to one another along common auditory dimensions e.g. amplitude and timbre, spatial dimensions e.g. proximity and orientation and embodied dimensions, e.g. weight and roughness. The balance method, illustrated in Figure 4.3, operates by modulating the relationships between the internal sonic elements of a soundscape in accordance with the data. This represents changes in the data source to a listener by altering their overall perception of balance within that soundscape.

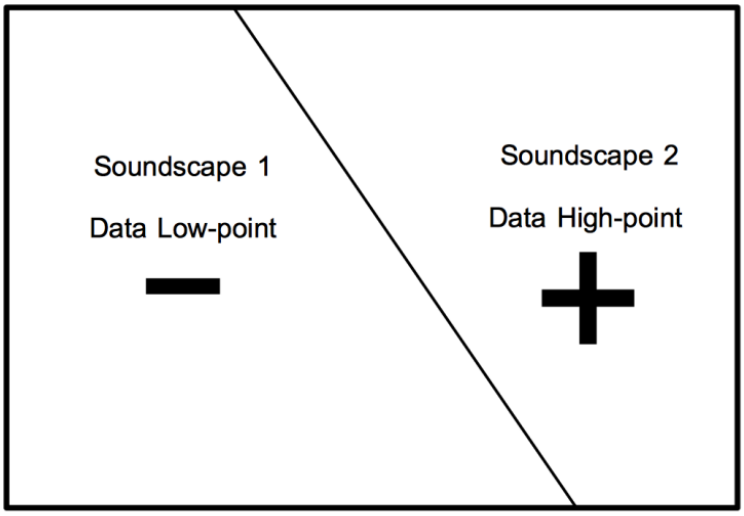


Figure 4.3. Balance

The second method of applying formal embodied structure is based on an embodied schema called the twin-pan balance schema proposed by Johnson (1987) as a special case of the balance schema. According to Johnson the twin-pan balance schema is a formal pattern that humans derive from their first-hand experiences of multiple counteracting forces finding balance around a fulcrum. The twin-pan balance method, illustrated in Figure 4.4, operates by selecting two distinct soundscapes, one that represents the high-point of the data, and one that represents the low-point and mixing between those soundscapes at a rate determined by the data set. This creates two distinct soundscapes for a listener at either extreme of a multivariate data set and a spectrum of transformations for all data-points in between.

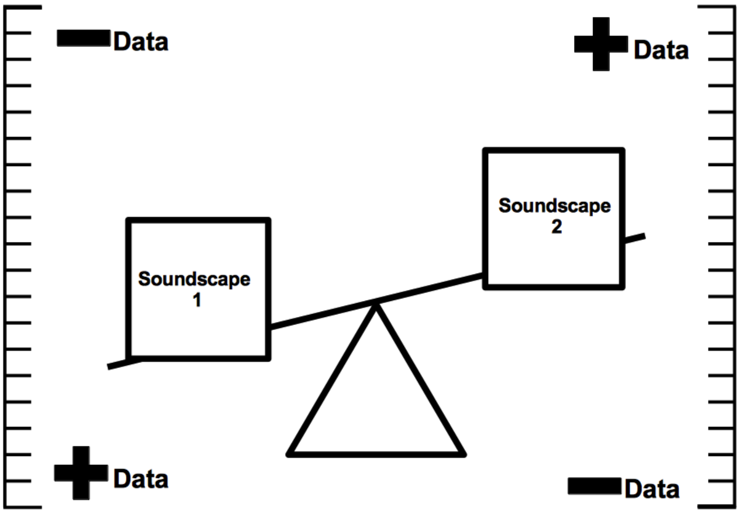


Figure 4.4. Twin-Pan Balance

4.6 Synthesising Candidate Frameworks

Both of the approaches to handling sonic complexes discussed here rely heavily on the communicative effectiveness of the sonic material used. When applying the framework it is crucial to use rich and communicative sonic material. Experimentation with prototyping platforms suggested that real world recordings provide better sonic complexes for soundscape sonification than synthesised sounds. It is also essential that each individual soundscape represents one single environment in order for the formal organisation methods to work effectively.

Metaphorically related sonic complexes should be chosen by either referencing the extensive body of research that exists on conceptual metaphors, or undertaking listener testing to determine new conceptual metaphors. Blended sonic complexes require that the data source has a causal effect on the sound source with the audible result representing the effect of the data source on the sound source within a shared location. The soundscape then represents this environment specific interaction.

Both of the approaches to organising the form of the soundscape sonification can be described by simple equations. The balance method operates by modulating the balance within a single soundscape. It is applied by establishing a base soundscape (Sl) that corresponds to the low point of the data and an extra set of sonic elements (Se) that, when fully integrated with the base soundscape create a transformed soundscape (Sh) that corresponds to the high-point of the data. The transformation from the base soundscape (Sl) to the transformed soundscape (Sh) is driven and controlled by the data series (D).

Both soundscapes (Sl and Sh) can be related to the data via the conceptual metaphor or conceptual blending techniques. This method is represented in Equation 4.1 where R1 and R2 represent the low and high points of the data-series' range and F(t) represents the state of the sonification at any given time, t. In the balance method the base soundscape (Sl) and transformed soundscape (Sh) must represent the same sonic environment, albeit somewhat transformed as the data increases.

$$F(t) = Sl + Sh(D/R2) + R1$$

Equation 4.1. Data-driven Soundscape Transformation

The twin-pan balance method operates by modulating the balance between two individual soundscapes. It requires the establishment of one soundscape that corresponds to the low point of the data (Sl) and another one that corresponds to the high point of the data (Sh). The high-point soundscape (Sh) is modulated by the data (D) and the low soundscape (Sl) is inversely modulated by the data (-D). This method is represented in Equation 4.2 below where R1 and R2 represent the low and high points of the data-series' range. F(t) represents the state of the sonification at an given time, t.

$$F(t) = Sl(-D/R2) + Sh(D/R2) + R1$$

Equation 4.2. Twin-Pan Data-Driven Soundscape Transformation

The design for each framework is quite simple and can be expressed in any number of audio processing languages and applications. The examples presented in this chapter were realised using Logic Pro X and Csound. A simple Csound implementation for applying both of these formal organisational strategies is included in the written appendices for this chapter and is also included in the digital appendices for this chapter on the appendices DVD.

4.7 Example Application of the Metaphorical Balance Method

The metaphorical balance method was applied to sonify the World Bank's figures for Irish GDP % growth rate from 1979 to 2013[16]. The entire data set is included in the digital appendices for this chapter on the appendices DVD, and graphed in Figure 4.5 below.

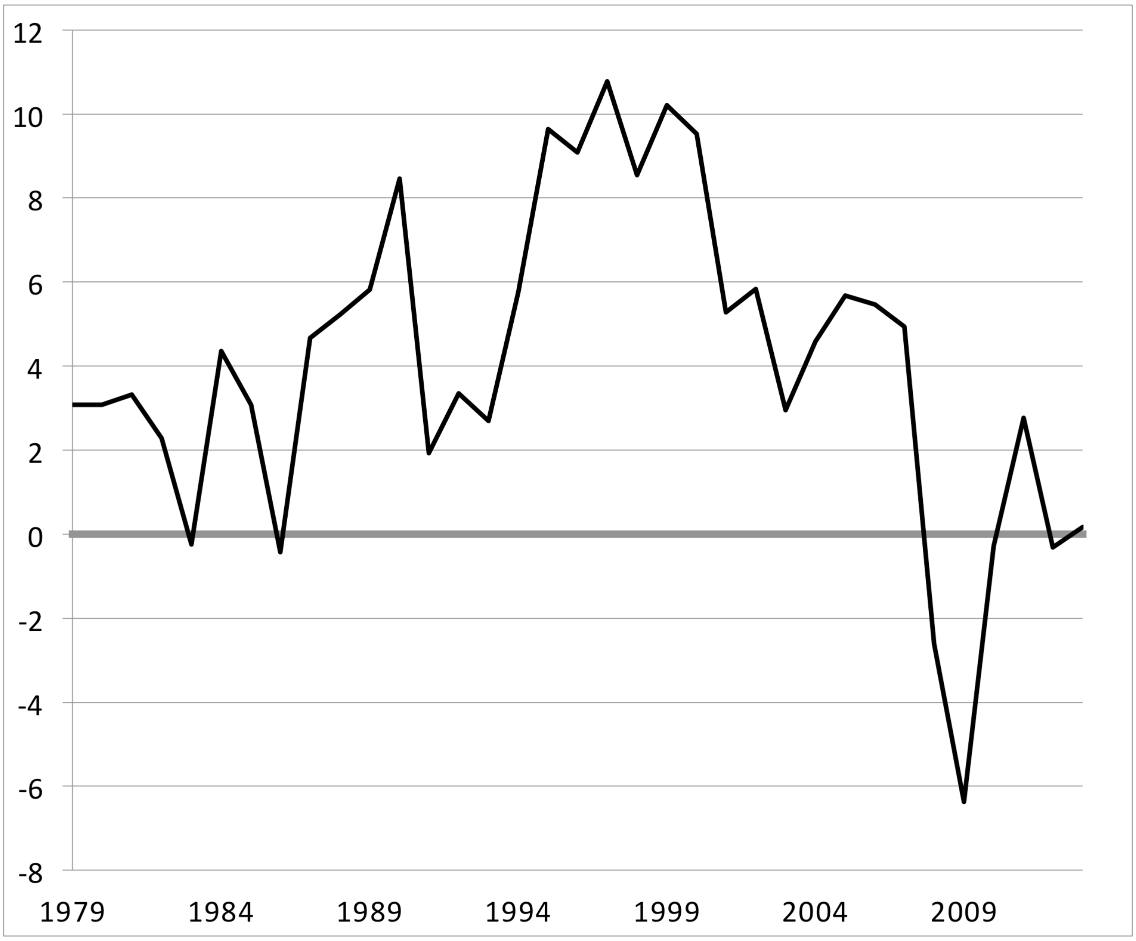


Figure 4.5. Irish % Annual GDP Growth from 1979 to 2013

The soundscape elements were retrieved from a number of sources including, most prominently, https://www.freesound.org/ and the BBC Sound Effects Library. More details on sound file attribution are included in the written appendices for this chapter and in the digital appendices for this chapter on the appendices DVD.

Two individual soundscape elements were used in the sonification. The base piece was intended to evoke the soundscape one would expect while experiencing smooth sailing on an old wooden ship. It is composed from a number of recordings of creaking wood and rope, boats gently breaking through the waves, and the calls of seabirds. The

---

[16] http://databank.worldbank.org/data/home.aspx

second element is composed from recordings of rough and stormy weather to evoke the soundscape of a storm[17]. This is a metaphorical soundscape sonification that draws on the conceptual metaphorical mapping where hard times and problems are reasoned about and understood in terms of bad weather (Żołnowska, 2011). This soundscape was created using Logic Pro X to arrange and compose the soundscape elements and Csound to format and map the data, to midi which was in turn ported back to Logic Pro X to control the amplitude balance between the two soundscape elements. The sonification is included in the digital appendices for this chapter on the appendices DVD.

4.8 Experimental Listener Evaluation: Selecting a Potential Candidate Framework

Two methods for selecting or designing sonic complexes and two methods of formal organisation of a soundscape sonification are laid out above. Drawing from these there are four possible frameworks for embodied soundscape sonification proposed. A listener evaluation of all four frameworks was undertaken in order to determine which one is the most communicatively effective. The most communicatively effective framework will be considered to be the one that expresses:

(A) The dynamic data curve of an entire sonification.

(B) The value of a static data-level presented in a sonification.

For the purposes of this chapter the term data curve is defined as a curve that links a series of successive data levels. Data curves change and evolve over time and are analogous to the curves used in visual trend graphs *etc.* The term data level is here used to refer to one of three levels of measurement, high, medium or low, represented in a visual graph or sonification. A static data level is one that does not change or evolve over the course of a sonification.

A listener evaluation was designed to test the effectiveness of each technique proposed in the soundscape sonification framework. Listeners were recruited from a large international pool of 40 countries via online crowdsourcing platform Crowdflower.com. This was done to ensure that the results were not specific to a particular culture and could

---

[17] A similar process of mapping environmental data to control sound object models derived from soundscape recordings used in Taylor and Fernström's *Marbh Chrios* (Taylor and Fernström, 2012).

be generalised across a large number of people. Experimental materials were hosted on Survey Gizmo. Participants were compensated for their participation. 128 people took part in the evaluations. Of that number 37 were female and 91 were male. 2 participants were under the age of 18. 19 participants were between 18 and 24. 69 participants were between the ages of 25 and 34. 32 participants were between the ages of 35 and 54 and 6 participants were over 55. Precautionary measures were taken to ensure that participants were using proper equipment. Listeners who could not pass a simple evaluation proving that they were using a 2-channel stereo system were barred from partaking in the evaluation. Listeners completed a short training trial to familiarise themselves with the procedure at the outset of the evaluation. The experiment took approximately 25 minutes to complete and aimed to answer questions A and B above.

4.9 Stimuli and Procedure

Listeners were presented with seven individual evaluations termed A to G. Six of these, A to F, were soundscape evaluations and G was a baseline evaluation intended to generate a reference point against which the results of A to F could be compared. This evaluation made use of pitch mapped sine tone stimulus to represent static and dynamic data values. The listeners completed two kinds of task during each evaluation, a matching task and an identification task. The matching task is intended to determine (A) which candidate framework can communicate the dynamic data curve of an entire sonification most effectively. It asked listeners to match visual representations of data curves with sonic representations of data curves created using the candidate embodied soundscape sonification frameworks. These dynamic stimuli fluctuate across three data levels, small, medium and large, over the course of a single presentation. They were 30 seconds long and formatted on the basis of each technique described in the embodied soundscape sonification framework. If listeners can correctly match a visual stimulus with the correct sonic stimulus they were said to have correctly identified the data curve from the sonic stimulus. Asking the listeners to draw the data curve may also have produced valid results. However this solution proved difficult to implement as the evaluations were carried out online.

The identification task (B) is intended to determine which candidate framework is most effective at communicating the individual data levels presented in a sonification. The identification task asked listeners to identify the values of static data levels presented using the candidate embodied soundscape frameworks. These static stimuli remain at a single data level small, medium or large, over the course of a stimulus. They were 6 seconds long and formatted on the basis of each candidate technique considered for the embodied soundscape sonification framework. Listeners were told that the stimuli presented in each of the evaluations represent one of the following socioeconomic data sets: unemployment rates, emigration rates, crime rates, GNP rates, suicide rates or poverty rates.

The stimuli used in this evaluation were purely auditory. They were not mapped to any actual data. They were designed to undergo specific transformations and occupy specific states in order to determine how listeners relate those transformations and states to an imagined data source. Accordingly the results obtained form the evaluation are not representative of any structure inherent in, or particular to, a specific data set, but rather describe how listeners map such sounds to imagined data values.

More details about the stimuli along with some sample evaluation questions are presented in the written appendix for this chapter. The stimuli, the entire set of evaluation questions and the raw data obtained during this evaluation are all included in the digital appendices for this chapter on the appendices DVD.

4.10 Statistical Analysis and Interpretation of Results

Content	**Structure**	**Evaluation**	**Observed number**			
Metaphorical						
			1	**2**	**3**	**4**
	Twin-pan					
		A	**81**	20	11	16
		E	52	**33**	15	28
	Balance					
		B	31	36	**36**	25
		C	34	18	**63**	13
		D	42	32	10	**44**
		F	**57**	30	21	20
Blended						
	Twin-pan					
		A	39	17	24	**48**
		D	**59**	29	14	26
		E	31	34	**31**	32
		F	40	29	40	**19**
	Balance					
		B	64	34	28	**13**
		C	**57**	14	13	44
Sine						
	Pitch Mapping					
		G	28	23	**62**	15

Table 4.1 *Listener responses for all stimuli. Correct choices in bold*

The matching task (A) was intended to determine how well listeners could determine the overall shape of the data curve from a soundscape sonification. The results for the matching task are presented in Table 4.1. Results can be interpreted against the baseline results generated for the pitch mapped sine tone. It should be noted that only 48% of listeners identified the pitch mapped sine tone correctly. This can be explained by the profile of the listeners tested in this evaluation. This thesis examines embodied cognition as a method for sonification design in the context of an increasingly networked sonification listenership. To reflect this participants in each evaluation represent a relatively large international pool of networked online listeners. The results of these evaluations describe sonification listening in the context of those conditions and so produce different results to those that would be expected from laboratory conditions. This is discussed in greater detail in section 2.4.1 of Chapter 2.

The strongest individual result of 81 correct identifications was recorded for the metaphorical twin-pan balance approach of evaluation A. This is 15% more than the 62 listeners who identified the pitch mapped sine tone correctly. The weakest individual result of 13 correct identifications was recorded for the blended balance approach of evaluation B. This is 38% less than the 62 listeners who identified the pitch mapped sine tone correctly. On average of 57 listeners identified the metaphorical twin-pan balance stimuli correctly, 55 listeners identified the metaphorical balance stimuli correctly, 39 listeners identified the blended twin-pan balance stimuli correctly and 35 listeners identified the blended balance stimuli correctly.

Chi-square goodness-of-fit tests (see Lancaster, 1969) were performed on the observed results of each matching task as represented in Table 4.1. These were intended to determine whether the results were statistically or practically significant.

Evaluation G examined the pitch mapped sine tone stimulus intended as a baseline against which the other stimuli could be compared. As discussed previously only 62 listeners correctly matched this stimulus to its visual counterpart X^2 (3, N = 128)=45.44, $p<.001$, V=.34. The significant *p* value and large effect size suggest reliable result with a large degree of practical significance.

Evaluations A and E examined metaphorical twin-pan balance stimuli. 81 listeners associated the correct visual data-contours with sonified data-contours for the metaphorical twin-pan balance stimulus A X^2 (3, N = 128)=101.313, $p<.001$, V=.51. The significant *p* value and large effect size represented by V suggest that a clear majority of listeners were easily able to identify the data-contour for this sonification by matching it to a visual representation. The results obtained for this evaluation are stronger than those recorded for the pitch mapped sine tone baseline. This suggests that the soundscape sonification stimulus examined in this evaluation might be better at expressing a data contour than the standard pitch mapping approach under the correct conditions. The significant *p* value and large effect size also suggest that the results are reliable and of a large degree of practical significance. 33 listeners associated the correct visual data-contours with sonified data-contours for the metaphorical twin-pan balance stimulus E X^2 (3, N = 128)=22.06, $p<.001$, V=.24. A clear majority of 52 listeners associated the sonification with the wrong visually graphed data contour. This suggests that listeners were confused by, the stimulus. The significant *p* value and medium effect size indicate a reliable result with a medium degree of practical significance.

Evaluations B, C, D and F examined metaphorical balance stimuli. The results for the metaphorical balance stimulus of evaluation B clustered around the 25% mark for each visual data contour X^2 (3, N = 128)=2.56, *p*>.05, V=.08. This indicated that listeners were confused by the task. The non-significant *p* value and small effect size further support this. 63 listeners, roughly 50%, associated the correct visual data-contours with sonified data-contours for the metaphorical balance stimulus of evaluation C X^2 (3, N = 128)=47.56, *p*<.001, V=.35. Because there were four possible matching options, the chance a listener would randomly select the right one is only 25% This suggests a strong result. The significant *p* value and large effect size suggest that the results are reliable and of a large level of practical significance. 44 listeners associated the correct visual data-contours with sonified data-contours for the metaphorical balance stimulus of evaluation D X^2 (3, N = 128)=22.75, *p*<.001, V=.24. This was not a significant overall majority. 42 listeners associated the stimulus with a different graph suggesting a measure of confusion for the listeners between these two options. The significant *p* value and large effect size suggest that this results is reliable and has a large degree of practical significance. 57 listeners associated the correct visual data-contours with sonified data-contours for the metaphorical balance stimulus of evaluation F X^2 (3, N = 128)=27.94, *p*<.001, V=.27. This result is slightly lower than the result of 62 recorded for the pitch mapped sine tone baseline suggesting that the stimulus is slightly less effective than a pitch mapped sine tone. The significant *p* value and medium effect size suggest that this is a reliable result of medium practical significance.

Evaluations A, D, E and F examined blended twin-pan balance stimuli. 48 listeners associated the correct visual data-contours with sonified data-contours for the blended twin-pan balance stimulus of evaluation A X^2 (3, N = 128)=18.56, *p*<.001, V=.22. The recorded figure is significantly lower than the result for the pitch mapped sine tone baseline. The significant *p* value and medium effect size suggest that this is a robust result of medium practical significance. 59 listeners associated the correct visual data-contours with sonified data-contours for the blended twin-pan balance stimulus of evaluation D X^2 (3, N = 128)=34.31, *p*<.001, V=.3. This result is slightly below the result for the pitch mapped sine tone baseline suggesting that the stimulus is slightly less effective than a pitch mapped sine tone. The significant *p* value and large effect size suggest that it is a reliable result of medium practical significance. The results are distributed almost evenly across all of the possible answers for evaluation E's blended twin-pan balance stimulus X^2 (3, N = 128)=.19, $p>.05$, V=.02. This suggests that they were due to chance and that listeners were confused by the stimulus. This is further supported by the non-significant *p* value and small effect size produced by the analysis. Only 19 listeners associated the correct visual data-contours with sonified data-contours for the blended twin-pan balance stimulus of evaluation F X^2 (3, N = 128)=9.56, $p<.001$, V=.16. The significant *p* value but small effect size produced by the analysis suggest that while these results are of practical significance a clear majority of listeners misunderstood the sonification.

Evaluations B and C examined blended balance stimuli. Only 13 listeners matched the soundscape sonification stimulus to the visually graphed data contour in evaluation B and a clear majority matched the sonic stimulus to the wrong visual graph X^2 (3, N = 128)=25.68, $p<.001$, V=.26. The results suggest that listeners were confused by the stimulus and the significant *p* value and medium effect size indicate that the result is reliable and of a medium level of practical significance. 57 listeners matched the sonic stimulus to the correct visual graph for the blended balance stimulus of evaluation C X^2 (3, N = 128)=45.44, $p<.001$, V=.34. This is only slightly below the result for the pitch mapped sine tone baseline suggesting that the stimulus is slightly less effective than a pitch mapped sine tone. The significant *p* value and large effect size of the analysis suggest that the result is reliable and has a large level of practical significance.

Content	**Structure**	**Evaluation**	**Observation**		
Metaphorical					
			Small	**Medium**	**Large**
	Twin-pan				
		A 1	**52**	35	41
		A 2	13	27	**88**
		A 3	9	**64**	55
		E 1	**41**	67	20
		E 2	26	**82**	20
		E 3	33	53	**42**
	Balance				
		B 1	7	30	**91**
		B 2	**49**	58	21
		B 3	17	**71**	40
		C 1	**55**	53	20
		C 2	41	**71**	16
		C 3	12	49	**67**
		D 1	28	51	**49**
		D 2	**28**	65	35
		D 3	24	**61**	43
		F 1	13	42	**73**
		F 2	25	**83**	20
		F 3	**63**	49	16

Table 4.2 *Listener responses for metaphor stimuli. Correct choices in bold*

Content	**Structure**	**Evaluation**	**Observation**		
Blended					
			Small	**Medium**	**Large**
	Twin-pan				
		A 1	28	**80**	20
		A 2	29	55	**44**
		A 3	**31**	59	38
		D 1	**40**	67	21
		D 2	17	65	**46**
		D 3	18	**90**	20
		E 1	32	41	**55**
		E 2	11	**62**	55
		E 3	**25**	42	61
		F 1	18	**68**	42
		F 2	13	56	**59**
		F 3	**42**	61	25
	Balance				
		B 1	35	61	**32**
		B 2	28	**72**	28
		B 3	**59**	39	30
		C 1	**40**	59	29
		C 2	13	**76**	29
		C 3	12	26	**90**

Table 4.3 *Listener responses for blended stimuli. Correct choices in bold*

Content	**Structure**	**Evaluation**	**Observation**		
Sine					
			Small	**Medium**	**Large**
	Pitch Mapping				
		G 1	16	42	**70**
		G 2	10	**78**	40
		G 3	**44**	57	27

Table 4.4 *Listener responses for pitch-mapped sine baseline. Correct choices in bold*

The identification task (B) was intended to determine how well listeners could determine the individual static data levels represented in a soundscape sonification. The results for the matching task are presented in Tables 4.2, 4.3, and 4.4. 55% of listeners identified the pitch mapped sine tone correctly. The strongest individual result of 91 correct identifications was recorded for a large stimulus created using the metaphorical balance approach in evaluation B. This is 16% more than the 70 listeners who identified the large stimulus of the pitch mapped sine tone correctly. The weakest individual result of 25 correct identifications was recorded for the small stimulus of the blended twin-pan balance approach in evaluation E. This is 14% less than the 70 listeners who correctly identified the corresponding small stimulus of the pitch mapped sine tone. This is 38% less than the 62 listeners who identified the pitch mapped sine tone correctly. On average 62 listeners identified the metaphorical twin-pan balance stimuli correctly, 63 listeners identified the metaphorical balance stimuli correctly, 53 listeners identified the blended twin-pan balance stimuli correctly and 61 listeners identified blended balance stimuli correctly.

A number of chi-square goodness of fit tests were carried out in order to further analyse the data. Evaluation G examined the pitch mapped sine tone baseline stimuli. The results for G provide the baseline against which other results can be interpreted. For evaluation G, 70 listeners identified the large value representation correctly X^2 (2, N = 128)=34.18, *p*<.001, V=.73. The significant *p* value and large effect size suggest a reliable result with a large level of practical significance. 78 listeners identified the medium stimulus correctly X^2 (2, N = 128)=54.44, *p*<.001, V=.92 the significant *p* value and large effect size suggest a reliable result with a large level of practical significance. Only 44 listeners identified the small value stimulus correctly X^2 (2, N = 128)=10.6, *p*<.01, V=.4. A clear majority of listeners incorrectly identified the stimulus suggesting that listeners were confused by it. The significant *p* value and large effect size suggest a reliable result with a large degree of practical significance.

Evaluations A and E examined metaphorical twin-pan balance Stimuli. For evaluation A, 52 listeners identified the small value stimulus correctly X^2 (2, N = 128)=3.48, *p*>.05, V=.12. The non-significant *p* value indicated that this result was unreliable. 64 listeners matched the medium X^2 (2, N = 128)=74.55, *p*<.001, V=.54 the significant *p* value and large effect size suggest a reliable result with a large degree of practical significance. 88 listeners identified the large value stimulus correctly X^2 (2, N = 128)=40.8, *p*<.001, V=.4. This result is larger than the result obtained for the baseline of evaluation G and the significant *p* value and large effect size suggest that this is a reliable result with a large degree of practical significance. For evaluation E only 41 listeners identified the small value stimulus correctly X^2 (2, N = 128)=25.98, *p*<.001, V=.32. A clear majority identified stimulus incorrectly. This suggests that listeners were confused by the stimulus. The significant *p* value and medium effect size suggest a reliable result with a medium degree of practical significance. 82 listeners identified the medium value stimulus correctly X^2 (2, N = 128)=54.81, $p<.001$, V=.46. This was a substantially larger result than that obtained for the pitch-mapped baseline and the significant *p* value and large effect size suggest a reliable result with a large practical significance. For the large value stimulus the results cluster around 33%, the number of observations expected from chance. X^2 (2, N = 128)=4.7, $p>.05$, V=.14. The analysis reveals that the results obtained are not statistically significant. This suggests that listeners were confused by the stimulus.

Evaluations B, C, D and F examined metaphorical balance stimuli. 91 listeners identified the large value stimulus correctly for evaluation B X^2 (2, N = 128)=88.3, $p<.001$, V=.59. This is a substantially higher result than that recorded for the sine baseline. The significant *p* value and large effect size indicate a reliable result with a large degree of practical significance. Only 49 listeners identified the small value stimulus correctly for B X^2 (2, N = 128)=17.45, $p<.001$, V=.26. A clear majority identified it incorrectly. The significant *p* value and medium effect size suggest that this is a reliable result with a medium degree of practical significance. This suggests that listeners were confused by this stimulus. 71 listeners identified the medium level stimulus correctly for evaluation B X^2 (2, N = 128)=33.42, $p<.001$, V=.37. This is close to but below the level established by the baseline suggesting that the stimulus is slightly less effective than a pitch mapped sine tone. The significant *p* value and medium effect size suggest a reliable result with a medium degree of practical significance. Only 55 listeners identified the small value stimulus correctly for evaluation C X^2 (2, N = 128)=18.11, $p<.001$, V=.27. 53 listeners incorrectly categorised the stimulus as medium, suggesting that listeners were confused by the stimulus. This was supported by the significant *p* value and medium effect size, which indicates a reliable with a medium level of practical significance. 71 listeners identified the medium value stimulus correctly for C X^2 (2, N = 128)=35.55, $p<.001$, V=.37. This is close to but below the level established by the baseline suggesting that the stimulus is slightly less effective than a pitch mapped sine tone. The significant *p* value and medium effect size suggest a reliable result with a medium level of practical significance. 67 listeners identified the large level stimulus correctly for evaluation C X^2 (2, N = 128)=36.86, $p<.001$, V=.38. This is only slightly below the level established by the baseline suggesting that the stimulus is only slightly less effective than a pitch mapped sine tone. The significant *p* value and large effect size suggest a reliable result with a large level of practical significance.

Only 49 listeners identified the large value stimulus correctly for evaluation D X^2 (2, N = 128)=7.61, $p<.05$, V=.17. 51 listeners incorrectly categorised the stimulus as medium, suggesting that listeners were confused by the stimulus. The significant *p* value and small effect size indicate a reliable result with a small degree of practical significance. Only 28 listeners identified the small value stimulus correctly for D X^2 (2, N = 128)=18.11, *p*<.001, V=.27. A clear majority identified the stimulus incorrectly. The significant *p* value and medium effect size suggest that this is a reliable result with a medium level of practical significance. 61 listeners identified the medium level stimulus correctly for evaluation D X^2 (2, N = 128)=36.86, *p*<.001, V=.38. This is below the level established by the baseline suggesting that the stimulus is less effective than a pitch mapped sine tone. The significant *p* value and large effect size suggest a reliable result with a large level of practical significance.

73 listeners identified the large value stimulus correctly for evaluation F X^2 (2, N = 128)=42.20, *p*<.001, V=.41. This is slightly above the baseline result and the significant *p* value and large effect size indicate a reliable result with a large level of practical significance. 83 listeners identified the medium value stimulus correctly for evaluation F X^2 (2, N = 128)=57.48, *p*<.001, V=.47. This is above the baseline result and the significant *p* value and large effect size suggest that it is a reliable result with a large level of practical significance. 63 listeners identified the medium value stimulus correctly for evaluation F X^2 (2, N = 128)=27.3, *p*<.001, V=.33. This is above the baseline result and the significant *p* value and medium effect size suggest that it is a reliable result with a medium level of practical significance.

Evaluations A, D, E and F examined blended twin-pan balance stimuli. For evaluation A, 80 listeners identified the medium value stimulus correctly X^2 (2, N = 128)=49.75, *p*<.001, V=.44. This result is substantially higher than the corresponding baseline result. The significant *p* value and large effect size indicate a reliable result of large practical significance. Only 44 listeners identified the large stimulus for evaluation A correctly X^2 (2, 128)=7.98, *p*<.05, V=.18. A clear majority incorrectly identified the stimulus. This indicates that listeners were confused by the stimulus. The significant *p* value and small effect size indicate that the result is reliable but has a small level of practical significance. Only 31 listeners identified the small stimulus correctly for evaluation A X^2 (2, 128)=9.95, *p*<.001, V=.2 and a clear majority of listeners incorrectly identified the stimulus. This indicates that listeners were confused by the stimulus. The significant *p* value and roughly medium effect size indicate a reliable result of medium practical significance.

For evaluation D, 40 listeners identified the small value stimulus correctly X^2 (2, N

= 128)=25.05, *p*<.001, V=.31. This result is slightly lower than the corresponding baseline result. The significant *p* value and medium effect size indicate a reliable result of medium practical significance. Only 46 listeners identified the large stimulus for evaluation D correctly X^2 (2, 128)=27.39, *p*<.001, V=.33. A clear majority incorrectly identified the stimulus. This indicates that listeners were confused by the stimulus. The significant *p* value and medium effect size indicate that the result is reliable and of medium practical significance. 90 listeners identified the medium stimulus correctly for evaluation D X^2 (2, 128)=78.81, *p*<.001, V=.55 The significant *p* value and large effect size indicate a reliable result with a high level of practical significance.

For evaluation E, 55 listeners identified the large value stimulus correctly X^2 (2, N = 128)=6.30, *p*<.05, V=.16. This result is lower than the corresponding baseline result. The significant *p* value and small effect size indicate a reliable result of limited practical significance. 62 listeners identified the medium stimulus for evaluation E correctly X^2 (2, 128)=35.83, *p*<.001, V=.37. This result is lower than the corresponding baseline result. The significant *p* value and large effect size indicate that the result is reliable and of large practical significance. Only 25 listeners identified the small stimulus correctly for evaluation E X^2 (2, 128)=15.20, *p*<.001, V=.24. A clear majority of listeners misidentified the stimulus, suggesting that they were confused by the stimulus. The significant *p* value and medium effect size indicate a reliable result with a medium level of practical significance.

For evaluation F, 68 listeners identified the medium value stimulus correctly X^2 (2, N = 128)=29.31, *p*<.001, V=.34. This result is lower than the corresponding baseline result. The significant *p* value and roughly large effect size indicate a reliable result of medium to high practical significance. 59 listeners identified the large stimulus for evaluation E correctly X^2 (2, 128)=31.05, *p*<.001, V=.35. 56 listeners categorised the stimulus as medium in size. This suggests that the stimulus was confusing. The significant *p* value and large effect size indicate that the result is reliable and of large practical significance. Only 42 listeners identified the small stimulus correctly for evaluation E X^2 (2, 128)=15.20, *p*<.001, V=.24. A clear majority of listeners misidentified the stimulus, suggesting that they were confused by the stimulus. The significant *p* value and medium effect size indicate a reliable result with a medium level of practical significance.

Evaluations B and C examined blended balance stimuli. For evaluation B, only 32 listeners identified the large value stimulus correctly X^2 (2, N = 128)=11.92, *p*<.01, V=.22. A clear majority of listeners identified the stimulus incorrectly. This suggests that the listeners were confused about the stimulus. The significant *p* value and medium effect size indicate a reliable result of medium practical significance. 72 listeners identified the medium stimulus for evaluation B correctly X^2 (2, 128)=30.25, *p*<.001, V=.34. This was lower than the corresponding baseline result. The significant *p* value and roughly large effect size indicate that the result is reliable and of medium to large practical significance. 59 listeners identified the small stimulus correctly for evaluation B X^2 (2, 128)=10.33, *p*<.01, V=.20. The significant *p* value and roughly medium effect size indicate a reliable result with a small to medium level of practical significance.

For evaluation C, only 40 listeners identified the small value stimulus correctly X^2 (2, N = 128)=10.80, *p*<.01, V=.21. A clear majority of listeners identified the stimulus incorrectly. This suggests that the listeners were confused about the stimulus. The significant *p* value and medium effect size indicate a reliable result of medium practical significance. 76 listeners identified the medium stimulus for evaluation C correctly X^2 (2, 128)=51.05, *p*<.001, V=.45. This was slightly lower than the corresponding baseline result. The significant *p* value and large effect size indicate that the result is reliable and of large practical significance. 90 listeners identified the large stimulus correctly for evaluation C X^2 (2, 128)=81.06, *p*<.001, V=.56. This is substantially larger than the figure recorded for the corresponding baseline. The significant *p* value and large effect size indicate a reliable result with a large level of practical significance.

4.11 Discussion of Results

Overall the results of these evaluations suggest that on average twin-pan balance soundscapes are more effective than balance soundscapes for communicating the shape of a data curve and are of roughly equal effectiveness for communicating individual discrete data levels to a listener. The results also suggest that the metaphorical soundscape stimuli were more communicatively effective than the blended soundscape stimuli for representing both a dynamically changing data curve and a static data level.

On average, none of the soundscape sonification techniques proved more effective than the pitch mapped sine tone baseline for communicating a data curve. However, the individual metaphorical twin-pan balance stimulus of evaluation A did prove more effective. 81 listeners identified this stimulus correctly and this result shows a high level of practical significance for sonification design as evidenced in the large effect size generated. This suggests that in general the metaphorical twin-pan balance approach holds the capacity to be more effective than the pitch mapped sine tone approach. However, further research is needed to determine why this stimulus produced such a strong results in comparison to the other metaphorical twin-pan balance stimulus that was correctly identified by only 33 listeners.

The results obtained in this evaluation are varied. Certain results were weak where a stronger result would be expected and other results were strong where a weak result would be expected. For example, all of the four blended twin-pan balance stimuli with the exception of E produced statistically and practically significant results. Similarly, the metaphorical twin-pan balance stimulus of evaluation A produced a much stronger result than that of evaluation E. Both stimuli use similar sonic content and embodied schematic organisational patterns as their counterparts. This suggests that decisions made during the design process represent an important factor in the creation of intelligible soundscape sonifications. Future work might determine specific design practices for creating communicatively effective embodied soundscape sonifications.

The results obtained for the metaphorical balance stimulus of matching task B and the blended twin-pan balance stimulus of evaluation E suggested a particular lack of communicative effectiveness. This may in part be due to the design of the stimuli. Amplitude modelling was used to control the balance between soundscapes for all of the stimuli in this evaluation. However, each of the soundscapes has it's own internal ebb and flow in terms of the amplitudes of individual sonic elements. As such the internal amplitude fluctuations of the stimuli may have interfered with the amplitude modelling applied to the soundscapes in the stimulus. This in turn would make it hard for listeners to match the stimulus to a corresponding visually graphed data curve. This highlights the difficulty of working with the complexities of real world soundscape recordings. This further underlines the importance of developing good design practices for working with soundscape materials in a sonification context.

In the identification task, listeners incorrectly identified the small stimulus of the pitch mapped sine baseline. They showed difficulties identifying all of the small blended twin-pan balance stimuli and one of two small blended balance stimuli. Listeners also showed difficulty identifying one of the two small metaphorical twin pan balance stimuli and three of the four small metaphorical balance stimuli. These results suggest that listeners find it difficult to determine low data levels for both pitch mapped sine tone stimuli and embodied soundscape sonification stimuli. On average soundscape sonification methods were less effective than the pitch mapping method. However, as discussed in the previous section a number of stimuli proved to be more effective than the small pitch mapped sine tone baseline. This suggests that in certain circumstances embodied soundscape sonifications can be more effective than pitch mapped sine methods. It may be possible that listeners consider the presence of any sound whatsoever to represent the presence of a positive data value. It may be better to use a mapping strategy where the lowest point in the data is represented by silence and lower data values are mapped to quieter soundscape amplitudes. Listeners were more successful at identifying medium and high value soundscape sonifications then low value soundscape stimuli. On average listeners were more successful at identifying data levels from the pitch mapped sine tone baseline than the soundscape sonification stimuli. However, a number of individual stimuli, discussed in the previous section, presented notable exceptions to this trend. This once again suggests that embodied soundscape sonification approaches have the capacity to be more communicatively effective than pitch mapped sonifications for expressing medium and high level data values. Further research is required to better understand listeners' interpretation of value levels from soundscape sonifications and to determine exactly when and why they are effective and how this effectiveness can be controlled.

While the results generated by this empirical listener evaluation are interesting they are preliminary in nature. Many of the results generated require further research that might more precisely pin point how embodied soundscape sonification can be used to maximise communicative effectiveness. Nevertheless, these results are of practical significance in the context of sonification design as evidenced in the strong effect sizes discussed in the previous section.

The results suggest that the metaphorical twin-pan balance method for creating embodied soundscape sonifications is more effective at communicating data curves and discrete data points to a listener. While the metaphorical balance approach provided some higher individual results for the identification task, the metaphorical twin-pan balance approach proved more robust overall. The results also suggest that this approach has the capacity to be more effective than pitch mapping sonifications under certain circumstances. The exact nature of these circumstances requires further research. A number of researchers have discussed a distinction between primary metaphors and complex metaphors (Grady, 1997; Lakoff and Johnson, 1999). Primary metaphors directly connect subjective or abstract domains to embodied experiences, for example the association people draw between the experience of warmth and the subjective impressions of affection. Complex metaphors are constructed from these primary metaphors through the process of conceptual blending, in similar fashion to how molecules are constructed from atoms through a process of chemical bonding. Primary metaphors are more immediately meaningful to a person because they are directly connected to embodied experience, and complex metaphors are less immediately meaningful because their connection to embodied experience is mediated by multiple primary metaphors the relationships between which must be grasped in order to understand these metaphors. The metaphorical soundscapes explored in this chapter are similar to primary metaphors because they relate directly to concrete embodied experiences. The blended soundscapes are analogous to complex metaphors because they reference multiple domains of embodied experience the relationships between which need to be unpacked and understood through a reversal of the conceptual blending process in order for the soundscape to become meaningful. This would add additional cognitive load to the listener and might explain why the metaphorical soundscapes performed better than the blended soundscapes.

It is likely that the twin-pan balance method creates soundscapes that can be more easily interpreted by the listener. In the balance soundscapes listeners need to pay close attention to the individual elements in the soundscape, as well as the overall state of the soundscape in order to get a sense of changes in the data. In the twin-pan balance soundscapes listeners only need to pay attention to which soundscape is more prominent in the mix and the degree to which they judge it to be prominent. This might explain why the twin-pan balance soundscapes performed better than the balance soundscapes.

The metaphorical method for relating data to sonic complexes and the twin-pan balance method for organising and mapping data to those sonic complexes represent an effective technique for the design of embodied soundscape sonifications. This indicates that these methods represent a viable framework for embodied soundscape sonification that might prove useful for sonification design and provide a focal point for further research. This framework is further defined in the following section.

4.12 The Chosen Embodied Soundscape Sonification Framework

The embodied soundscape framework is a framework for creating soundscape sonifications that exploits the embodied nature of human cognition as described in the ESLM to effectively communicate data to a listener.

In this framework two soundscapes are used to represent a single data set. The first bears a metaphorical relation to the high point of the data and the second bears a metaphorical relation to the low point of the data. The original data set is mapped to modulate the perceptual salience of each soundscape in an inversely proportional ratio.

The entire framework can be described by the following transfer functions where msLo is the metaphorically related soundscape that represents the low point of the data msHi is the metaphorically related soundscape that represents the high point of the data, D represents the Data, R1 and R2 represent the low and high points of the data-series' range respectively and F(t) is the state of sonification at a given time, t.

$$F(t) = msL(-D/R2) + msH(D/R2) + R1$$

Equation 4.3 The Embodied soundscape Mapping Transform

4.13 Limitations of the Framework

One limitation of the embodied soundscape framework is that it has only been empirically verified to work for dynamic data curves and static sonifications across three levels of data: low, medium and high. As such, the framework is in its current state is best suited to conveying an overall topological gist of a data set. Further research is required to determine how effectively the framework performs for high resolution data sets with multitudes of data levels. The framework's limited capacity for representing higher resolution data must be offset against the results obtained when the framework was compared to the pitch mapping of a sine tone. These results suggest that in the correct circumstances the framework can potentially produce more communicative data-sonifications than the pitch mapping approach across three data-levels.

Another limitation of this approach is that the four candidate frameworks utilised only two of the many possible embodied schematic approaches to structuring the form of a sonification: the balance and twin-pan balance schemata. An exploration of different kinds of embodied schematic structuring strategies may make for even more effective sonifications.

4.14 Discussion

This chapter aimed to further explore the concept of sonic complexes described by the ESLM. In order to achieve this, it adopted environmental soundscape as a domain of sonic complexes and attempted to create an effective soundscape sonification framework that operated by mapping data to organise those sonic complexes.

It tackles the question of how sonification could be used to make communicatively effective soundscape sonifications. It provides a solution to this question in the shape of the embodied soundscape framework. The three research methods deployed in Chapter 3 to investigate embodied sonic dimensions, exploration with prototyping platforms, research via the data-driven compositional process and empirical listener evaluation, were also used in this chapter to explore the concept of sonic complexes.

The key finding of the experimental listener evaluation is that sonifications made to the specifications of the embodied soundscape framework, have the potential to be more communicatively effective than pitch mapped sine tone sonifications, for both dynamic data-curves and statically presented data-levels. The mapping of data values to pitch is a prevalent practice across the field of sonification (Grond and Berger, 2011) making this a relevant finding. This supports the claims made in this thesis that communicatively effective sonifications can be created by exploiting the embodied nature of sonification listening.

These results provide strong support for ESLM's claims about the role of conceptual metaphor and embodied schematic knowledge in sonification listening as the chosen framework is designed to exploit the conceptual metaphorical mappings that underlie sonification listening and the listener's embodied schematic knowledge of the twin-pan balance schema.

This thesis argues that each individual sonic complex, in this case a soundscape, has its own set of intrinsic sonic dimensions. These are the same embodied sonic dimensions studied in the context of vocal stereo spatial gestures in Chapter 3. The embodied soundscape framework presented here does not map data to these intrinsic dimensions but rather maps data to extrinsic sonic dimensions that are used to organise relationships between multiple sonic complexes.

Idle Hands and *Doom & Gloom* illustrate how an embodied approach to sonification can be applied to aesthetic practice in the context of data-driven composition. These pieces were also critical to the development of the embodied soundscape framework. *Idle Hands* has been performed at a national level, as discussed earlier. *Doom & Gloom*, which follows a similar structure to that of the embodied soundscape framework, has been performed extensively throughout 2015 as discussed earlier in this chapter.

4.15 Conclusions

This chapter opened with a consideration of embodied soundscape sonification before discussing some prototyping platforms and exploratory procedures used to explore how an embodied approach to soundscape sonification might be developed. A number of prototyping platforms are then introduced and exploratory research undertaken with those platforms in order to guide the design of a framework for embodied soundscape sonification is described. At this point two data-driven musical compositions, *Idle Hands* and *Doom & Gloom*, are introduced and discussed. The compositional process was used as an exploratory research method for evaluating how data could be mapped to a soundscape in a way that exploited the embodied nature of sonification listening. This is discussed in greater detail in appendix 4.3. Drawing from the prototyping platforms and compositions four possible frameworks for embodied soundscape sonification were then introduced and an example of one of the frameworks was presented. These frameworks were submitted to user testing the results of which determined the most effective framework for embodied soundscape sonification to be that in which a sonic complex is metaphorically related to the data source and the soundscape is organised around the structure of Johnson's (1987) twin-pan balance schema.

Chapter 5: Embodied Sonic Metaphor - The Temporo-spatial Motion Framework.

5.1 Introduction

This chapter provides a solution to the problem of representing temporal context in the sonification of time-series data. It introduces some of the conceptual metaphors involved in the cognition of music and time. These metaphors are extended to the domain of sonification in the proposal of the TIME-SERIES DATA AS PHYSICAL MOTION[18] metaphor. The proposed temporo-spatial motion framework leverages this conceptual metaphor to solve the temporal context problem in sonification. The proposed framework is then submitted to rigorous empirical testing to determine how well it achieves this aim. The results of this testing are used to inform a refinement of both the framework and the underlying conceptual metaphor that motivates it. The refined metaphor and framework are presented and discussed at the end of the chapter.

5.2 Motivation for Temporal Context

The term temporal context is used in this thesis to reference both the listener's understanding of where they are in a sonification in regards to its start and end points and where they are in a time series in regards to what time-points they are currently hearing, have previously heard and will hear before the sonification ends. The representation of temporal context to a listener can be problematic in sonification. Visual graphing is not time bound. At any given time the entire time-series in a graph is accessible to an observer. Visual graphs present static information to an observer who can choose where in the graph to direct their attention. Auditory graphs evolve and unfold as time-bound experiences rather than static objects. They must dynamically present their information and as a result the listener does not have the luxury of directing her/his attention to sets of features that endure unchanging through time. This presents some issues for the sonification of time-series data where a sense of temporal context, the listener's position in the time-series and knowledge of previous and future time-points, can be key to interpreting a sonification or auditory display. When attending to a well designed visual graph of time-series data, the observer can clearly identify individual time-points and

---

[18] Capitalisation is used in cognitive science literature to denote a conceptual metaphor.

temporal relationships, on the basis of context cues like axis markings, and labels (Hermann, 2002, p. 12). Auditory context cues are also required for a listener to interpret an auditory graph (Walker and Nees, 2011). The framework described in this chapter offers an embodied solution to this problem of temporal context in sonification.

5.3 Embodied Cognitive Underpinnings of Time and Music

The proposed temporo-spatial motion framework borrows from the work of Cox (1999) and Johnson and Larson (2003), which will be discussed shortly. They in turn build upon Johnson's earlier work with Lakoff (Lakoff and Johnson, 1980) on the metaphorical conceptualisation of time. These researchers have shown that time is conceptualised metaphorically as movement through space. This means that people frame, understand and reason about the abstract concept of time in terms of the familiar concepts of space and spatial movement. For example, Johnson and Larson (2003) point out that words used to speak about the passage of time, e.g., passage, flying, crawling, coming, going, front, behind, are also used to talk about movement through space. Two chief systems of metaphor, MOVING TIMES and MOVING OBSERVER are thought to define how time is understood and reasoned about (Cox, 1999; Johnson and Larson, 2003; Lakoff and Johnson, 1980). According to Johnson and Larson (2003):

> In the *"MOVING TIMES"* metaphor, the times are the figure moving relative to the stationary observer (as ground), whereas in the *"MOVING OBSERVER"* metaphor, the observer is the figure moving relative to the time landscape (as ground).

In both systems future times approach the observer's location from the front, become present when they share a location with the observer, and fade into the past when they pass behind the observer's location. The basic cross-domain mappings involved in these metaphorical systems, as documented by Johnson and Larson (2003), are presented in Tables 5.1 and 5.2. Both systems are thought to play a key role in the observer's understanding of time.

Source Domain (Space)		**Target Domain (Time)**
Objects	=	Times
Motion of objects past the observer	=	The passage of time
Location of the observer	=	The present
Space in front of the observer	=	The future
Space behind the observer	=	The past

Table 5.1. *The Moving Times Metaphor*

Source Domain (Space)		**Target Domain (Time)**
Locations on observer's path	=	Times
Motion of the observer	=	The passage of time
Distance moved by the observer	=	Amount of time passed"
Location of the observer	=	The present
Space in front of the observer	=	The future
Space behind the observer	=	The past

Table 5.2. *The Moving Observer Metaphor*

Johnson and Larson (2003) extend this metaphorical conception of time as spatial movement to music, arguing that listeners understand and reason about music as if it were spatial movement also. This argument is based around the three core systems of metaphorical mappings of MOVING MUSIC, MUSICAL LANDSCAPE and MUSIC AS MOVING FORCE. These metaphors draw from the listeners' perceptual experience of objects in motion, self-determined motion through spatial fields, and forced motion through spatial fields. In MOVING MUSIC the music moves past a listener, in MUSICAL LANDSCAPE the listener navigates a path through a musical landscape. In MUSIC AS A MOVING FORCE the music takes hold of and moves the listener through a series of locations.

Sonification listening, as described in the ESLM, draws on the same embodied cognitive meaning-making faculties as music. It could naturally be argued then that elements of the three metaphors introduced above could be extended to account for sonification. Tables 5.3, 5.4 and 5.5 propose a possible set of mappings for these proposed MOVING SONIFICATION, SONIFICATION LANDSCAPE and SONIFICATION AS A MOVING FORCE metaphors.

Source Domain (Physical Motion)		**Target Domain (Sonification)**
Physical Object	=	Data Point
Physical Motion	=	Movement through Time Series
Speed of Motion	=	Speed of movement through data
Location of Observer	=	Location in the Data set
Objects in front of observer	=	Future data point
Objects behind observer	=	Past data point
Path of motion	=	Value of the data
Starting/ending point of motion	=	Beginning/end of data record section
Temporary cessation of motion	=	Plateau in the data
Motion over same path again	=	Repeat of data structure
Physical forces	=	Causes

Table 5.3. *The Proposed Moving Sonification Metaphor*

Source Domain (Physical Space)		**Target Domain (Sonification)**
Traveller	=	Listener
Path traversed	=	Sonification
Traveller’s present location	=	Current data point
Path already travelled	=	Data already heard
Path in front of traveller	=	Data not yet heard
Segments of the path	=	Segments of the data
Speed of traveller’s motion	=	Speed of playback

Table 5.4. *The Proposed Sonification as Landscape Metaphor*

Source Domain (Physical Space)		**Target Domain (Sonification)**
Locations	=	States of understanding
Movement (from place to place)	=	Growth of understanding
Physical forces	=	Causes
Forced movement	=	Causation
Intensity of force	=	Intensity of Sonification Impact

Table 5.5. *The Proposed Sonification as a Moving Force Metaphor*

5.4 Proposed Time-Series as Motion Metaphor

Taking these three metaphors as a guide, the hypothetical TIME-SERIES DATA AS PHYSICAL MOTION metaphor is proposed here. This metaphor is intended to describe how listeners might make sense of time-series data in a sonification by relating it to the domain of motion through spatial fields. The possible mappings of this proposed metaphor are outlined in Table 5.6[19]. It derives from the previously discussed metaphorical systems listed above. Taking this proposed metaphor as a basis, the task of finding a solution to the problem of representing temporal context to a listener in the sonification of time-series data can begin.

Source Domain (Physical Motion)		**Target Domain (Sonification)**
Physical object	=	Data Point
Change in physical object	=	Change in data value
Vertical path of motion	=	Value of the data
Movement past listener Front/Back	=	Passage of unit of time
Movement past listener Left/Right	=	Passage through data set
Listener's horizontal position	=	Midpoint of the data
Starting/ending point of motion Left/Right	=	Beginning/end of time series
Starting/ending point of motion Front/Back	=	Beginning/end of time unit
Temporary cessation of motion	=	Plateau in the data
Speed of Motion	=	Speed of movement through data

Table 5.6. *Proposed Time-Series Data as Motion Metaphor*

5.5 Primary Aims and Secondary Aims of the Framework

The proposed temporo-spatial motion framework is intended to give listeners a sense of temporal context during sonification listening by exploiting the mappings in the proposed TIME-SERIES DATA AS PHYSICAL MOTION metaphor. This temporal context works on two levels. It gives listeners a sense of temporal context within the

---

[19] Table 5.6 represents a conceptual metaphorical mapping strategy proposed. This is a cognitive mapping strategy proposed to exist in the mind of a sonification listener. It is intended to describe how a listener might interpret a sonification of time-series data in terms of physical motion. It does not represent a mapping strategy from data to sound.

sonification itself by offering cues to support the listener in estimating temporal distance from the start, end and mid-points of a sonification. It also gives listeners a sense of the passage of time on the level of the individual time-points within the time-series itself. The primary aims of this framework are listed below. Using this framework a listener should be able to effectively estimate the following points of information during the sonification listening process:

(A1) The value of a time-point to which they are listening.

(A2) Their temporal position in the time-series in relation to the beginning and end points.

Though the framework is focused on the accurate representation of the time-series, the question of data-values related to those time-points is still of critical importance. Flowers (2005) points out that sonifications containing short individual data streams that can be easily cross-compared and are presented in a sequential manner are more intelligible than continuously represented data-streams. The proposed framework will represent data in this manner, while attempting to link data values to time-points. As such, the secondary aims of the proposed framework are intended to account for the representation of data-values associated with the time-point values. Using this framework a listener should be able to effectively estimate the following points of information during the sonification listening process:

(B1) Which data-points relate to which time-points.

(B2) The relationships between sequential data-points presented in temporal sequence.

5.6 Proposed Framework Specification

A single stream of time-series data can be broken into two components, the observed data values, and the time values associated with those data values. Time values are routinely coded using a standardised unit of time (e.g., days, weeks, months, years *etc.*). The framework operates by mapping data to salient auditory parameters of the designer’s choice (e.g., pitch, timbre) to create sonic data objects. Over the course of a single unit of time, a data object will seem to "pass by" the listener, moving from a position in front of the listener to a position behind them. This trajectory, from front to back, represents the passage of a specific unit of time as illustrated in Figure 5.1 below. Individual units of time may represent a single data measurement, or evolving data-contours across a finer time scale within that unit.

These passing data objects are organised from left to right, relative to the listener's position, across the horizontal axis of the stereo spatial field. The earliest data object appears to the extreme distant left of the listener, the mid-point of the data set appears at the listener's location and the last data point appears to the extreme distant right of the listener. This configuration, represented in Figure 5.2, creates an evolving matrix of time indexed data-driven gestures that pass by the listener's position in sequential order moving from left to right across the stereo spatial field.

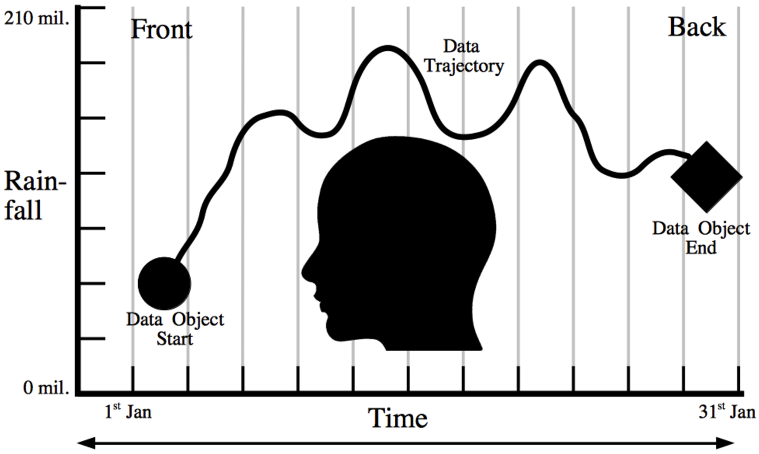


Figure 5.1. Unit Passage from Front to Back

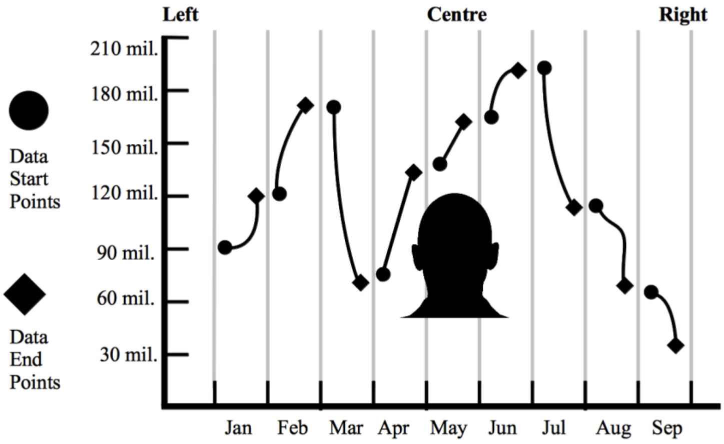


Figure 5.2. Time Series Passage from Left to Right

5.7. Synthesising the Framework

The framework can be synthesised and applied to sonify data by using a number of audio processing techniques currently used in the sound design field. Distance from the listener's position can be controlled through the use of reverb. Wet/Dry signal ratio (R.W/D), pre-delay time in ms and early reflection amplitude (E.R.Amp) can all be manipulated to create a perceptible sense of distance from the listener's position. Simple left-right panning can be used to distribute the data objects across the stereo-spatial field. The passage of data objects over units of time can be achieved by applying an algorithm to simulate the Doppler effect. The Doppler effect describes the phenomenon where a sound is perceived as higher in frequency when approaching a listener but lower in frequency when retreating from a listener. As a sound source approaches a listener, the gap between each successive sound wave becomes shorter, corresponding to a rise in frequency, and as it retreats the gaps between each wave lengthen, corresponding to a lower frequency. It is here argued that the listeners' identification and understanding of the Doppler shift, as indicative of a passing sound source, represents a thoroughly embodied structure of knowledge rooted in one's experience of physical motion through spatial fields. The task of representing front and back orientation in stereo-space is a complex one that is not yet solved. However, the Doppler effect leverages the listener's embodied knowledge of physical sound sources in motion to achieve this in a limited way for moving sound sources.

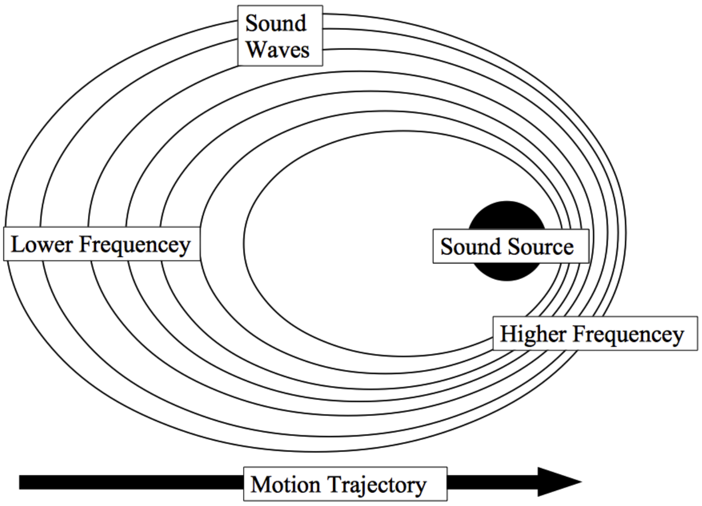


Figure 5.3. Doppler Effect

5.8. Application of the Framework

The framework was applied to organise a PMSon, which mapped monthly rainfall measured at Dublin Airport over 2014 to the pitch of a sine tone. PMSon was chosen over the more embodied approaches discussed in this thesis for a number of reasons. The first reason was to demonstrate that in some cases embodied cognition organisational principles can be applied to standard sonification techniques to create new layers of meaning that might enhance a sonification. The second reason was to ensure that the results obtained in this chapter might be useful to the wide range of researchers and practitioners that employ pitch mapping as a sonification technique. The data set provided by MET Éireann[20] is presented in Table 5.7

---

[20] The data is available at http://www.met.ie/climate/monthly-data.asp?Num=532

Month	**Rainfall**
January	101.6
February	88.5
March	53.7
April	34.2
May	91.5
June	36.2
July	35
August	173
September	26.5
October	90.2
November	140.9
December	55.9

Table 5.7. *Rainfall in millimetres recorded at Dublin Airport over 2014*

Data objects were generated in Csound where the data set was mapped to control the pitch in a poscil opcode. The poscil opcode, a contraction of precise oscillator, offers fine-grained control of frequency values. It was used to synthesise a simple sine wave signal. Rainfall data was mapped to modulate the frequency of this sine wave across a range from 220Hz (A3) to 1760Hz (A5). Data objects are distributed across stereo-space by panning them between stereo sources and applying reverb modelling to simulate distance. The parameters used are presented in Table 5.7.

Each individual data object is nine seconds long and represents the change in value from one month to the next. For example, the first data object represents the change in value from 101.6mm of rain for January to 88.5mm for February, as mapped to pitch. A Doppler algorithm is applied to each data object to simulate a Doppler shift at the half-way point, 4.5 seconds, in the sound's playback. The algorithm also applies some more cues to simulate distance. It incrementally increases the gain of sound sources on approach and decreases them as they retreat. The algorithm also adds a small amount of reverb (10% of overall signal amplitude) to the signal, which it decreases upon approach and increases upon retreat.

Month	**Pan Degrees**	**Reverb Wet: Dry Ratio**	**Reverb Pre-delay Time in Milliseconds**	**Early Reflection Amplitude**
Jan	-90	0:100	100	-30
Feb	-75	20:80	80	-24
Mar	-60	40:60	60	-18
Apr	-45	60:40	40	-12
May	-30	80:20	20	-6
Jun	-15	100:00	0	0
Jul	15	100:00	0	0
Aug	30	80:20	20	6
Sep	45	60:40	40	12
Oct	60	40:60	60	18
Nov	75	20:80	80	24
Dec	90	0:100	100	30

Table 5.8. *Control Parameters for Left to Right Distribution*

5.9 Experimental Listener Evaluation

A listener evaluation was designed to test the effectiveness of the framework. Listeners were recruited from across a large international pool of 49 countries via online crowdsourcing platform Crowd Flower. This was done to ensure that the audience composition reflected a cross section that of a relatively large international group of non-specific listeners listening online. Experimental materials were hosted on Survey Gizmo. Participants were compensated financially for their participation. 191 people took part in the evaluations. Of that number 72% were male and 28% were female. 18% of participants were between 18 and 24. 43% of participants were between the ages of 25 and 34. 35% of participants were between the ages of 35 and 54 and 3% of participants were over 55. Precautionary measures were taken to ensure that participants were using proper equipment. Listeners who could not pass a simple evaluation proving that they were using a 2-channel stereo system were barred from partaking in the evaluation. Listeners completed a short training trial to familiarise themselves with the procedure at the outset of the evaluation. The experiment took approximately 25 minutes to complete. The experiment aimed to test how effectively listeners can estimate:

(A1) The value of a *time-point* to which they are listening.

(A2) Their temporal position in the time-series in relation to the beginning and end points.

(B1) Which data-points relate to which time-points

(B2) The relationships between sequential data-points presented in temporal sequence.

Listeners were presented with five individual evaluations labelled A to E. Each evaluation had an associated audio stimulus. Listeners were allowed to replay the sonification stimuli as many times as was needed to answer the questions. Audio A contained the entire rainfall sonification and had a duration of 1 minute and 16 seconds. Audio B contained the sonified data for January, February and March and had a duration of 21 seconds. Audio C contained the sonified data from January to September and had a duration of 56 seconds. Audio D contained the sonified data for August and September and had a duration of 15 seconds. Audio E contained the sonified data for April and May and also had a duration of 15 seconds. Questions 1 and 2 in evaluation A were intended to determine how effectively listeners could estimate the data values associated with specific time-points. Questions 3 and 4 in evaluation B and questions 7 and 8 in evaluation C were intended to determine how well listeners could estimate their position in the data series. Questions 5 and 6 in evaluation B and questions 9 and 10 in evaluation C were intended to determine how well listeners could estimate their position in the data series. Questions 11 and 12 in evaluations D and E were intended to test how well listeners could cross compare sequentially presented data using this framework. Each of the questions in the evaluation provided a 'Don't Know' option allowing participants to report that they did not know the correct option.

5.10 Statistical Analysis and Interpretation of Results

The listener responses for each question were analysed using two chi-square goodness-of-fit tests. This analysis method was chosen to help determine whether or not the frequency distribution of listeners' responses deviated from the mean expected responses to a statistically significant degree. The first test is performed on the results for all of the options in the evaluations including the don't know options. This gives a sense of the overall statistical significance and practical significance, as indicated by the effect size measure: Cramer's V, of the data. The second test is performed on all of the results excluding the don't know options. Performing both of these tests allows for the identification of cases where statistical significance and effect size are caused by the number listeners who chose the ‘don’t know’ option, indicating that they did not understand the stimuli and/or question asked of them.

5.11 B2. How effectively can listeners link data points to time-points

For Q1 and Q2 listeners were presented with the entire sonification and asked to determine which months had the highest and lowest rainfall respectively. These results are illustrated in figures 5.4 and 5.5 respectively. The data used for these sonifications is presented in Table 5.7.

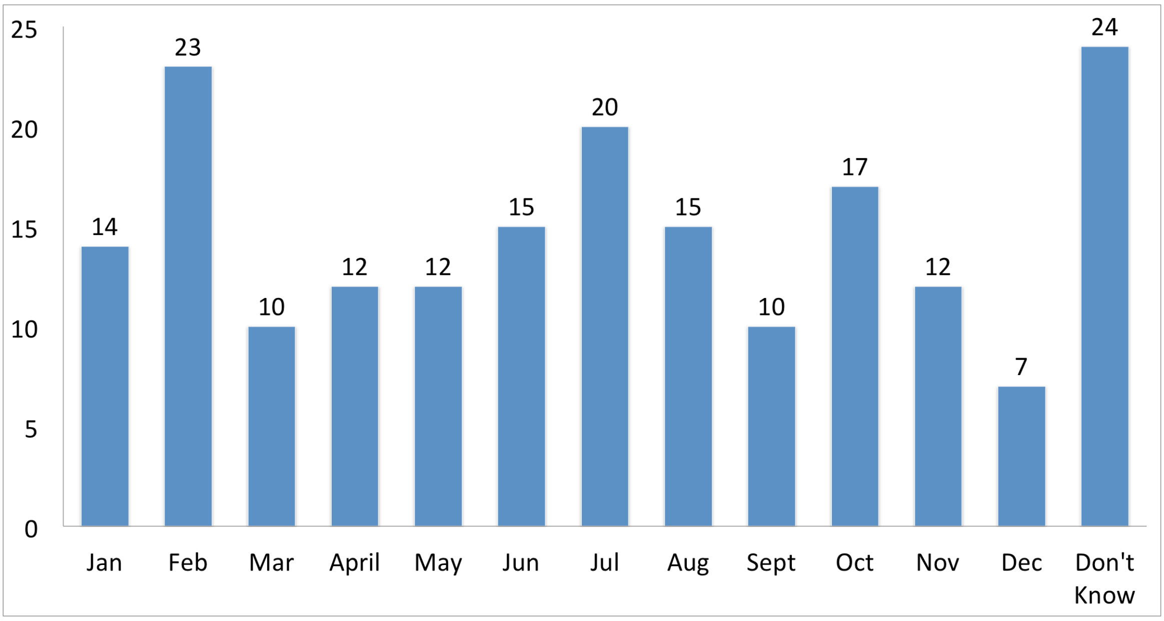


Figure 5.4. Data Breakdown for Question 1

The high point of the data was in August, but only 15 listeners identified this from the sonification. February was selected with the highest frequency of 23 listeners. This was followed by July with 20 listeners and then October with 17. This suggests that listeners could not effectively link data-values with time-points. 24 listeners selected the don’t know option. This was roughly the same number that selected the correct option.

Two chi square goodness-of-fit tests were carried out on the results. The first test included the results for each of the month options and the don’t know option X^2 (12, 191)=21.42, $p<.05$, V=.1. The second test considered only the results for the month options $X^2(11, 167)=15.87$, $p>.05$, V=.09. The first test produced a highly significant p-value of $p<.001$ but a small effect size of V=.1. This suggests that the results are of high statistical significance but little practical significance. The second test generated a non-significant p-value of $p>.05$ and a small effect size of V=.09. This further suggests that the number of listener's choosing the correct option over the incorrect alternatives was statistically non-significant and of almost no practical significance.

The results for question 1 suggest a substantial number of listeners misinterpreted the sonification by choosing incorrect options and a similarly substantial number of listeners reported that they did not understand the sonification. This indicates that listeners could not effectively link the highest data point to its time point using the sonification.

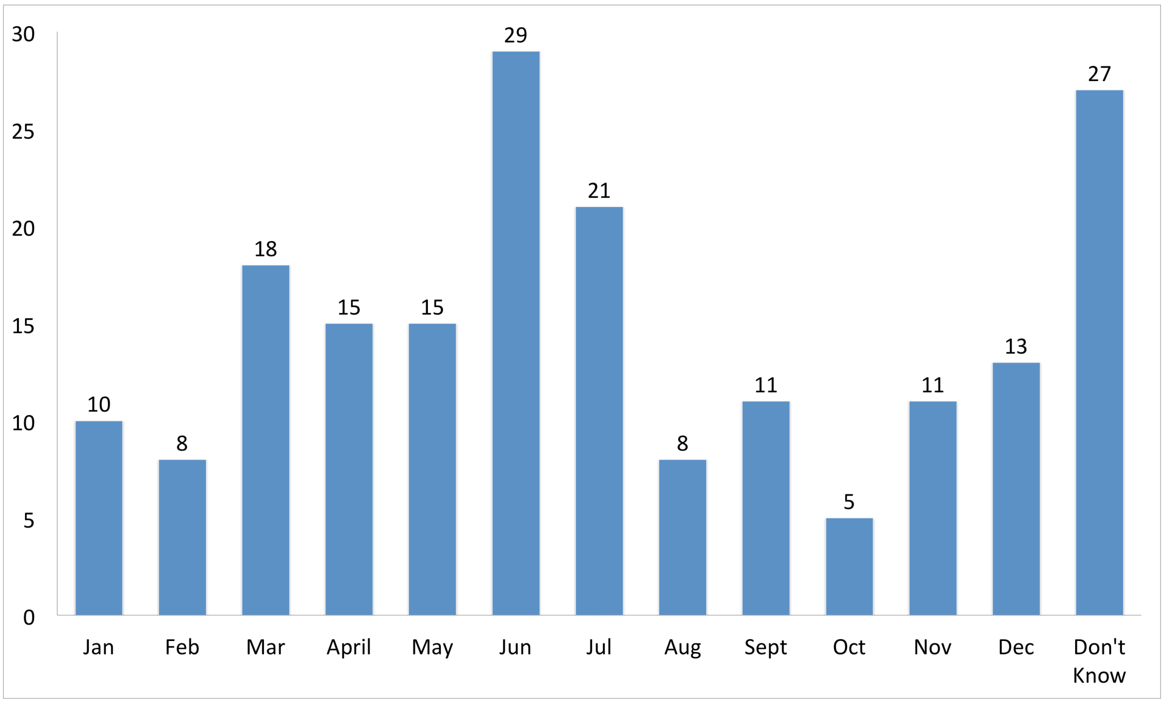


Figure 5.5. Data Breakdown for Question 2

In question two, the low point of the data was in April, but only 15 listeners identified this from the sonification. June was selected by the highest number of listeners 29, followed by July with 21 listeners and March with 18. Furthermore 27 listeners reported that they did not know the correct answer. This suggests that listeners could not effectively link data-values with time-points.

Two chi square goodness-of-fit tests were carried out on the results. The first test included the results for each of the month options and the don't know option $X^2(12, 191)=43.74$, $p<.001$, V=.14. The second test considered only the results for the twelve month options $X^2(11,164)=35.02$, $p<.001$, V=.14. The first test produced a highly significant p-value of $p<.001$ and a small effect size of V=.14. This suggests that the results are of high statistical significance but a small level of practical significance. The second test produced a significant p-value of $p<.001$ and a small effect size of V=.14.

This shows that the number of listener's choosing the incorrect options over the correct option was of high statistical significance but little practical significance.

The results for question 2 suggest a substantial number of listeners misinterpreted the sonification by choosing incorrect options and a similarly substantial number of listeners reported that they did not understand the sonification. This indicates that listeners could not effectively link the lowest data point to its time point using the sonification. The results for question 1 and 2 suggest that listeners cannot effectively link data values with time-points when listening to a sonification organised in terms of the temporo-spatial motion framework.

5.12 A1. What Time-Point they are Currently Hearing

For Q3 and Q7 listeners were presented with sonifications and asked to determine the half of the year in which the sonifications presented their final data-points. Listeners could select either half of the year or a 3rd choice of "Don't Know". For Q4 and Q8 listeners were presented with sonifications and asked to determine the exact month on which those sonifications ended. Listeners could select any of the 12 months or a 13th option of "Don't Know".

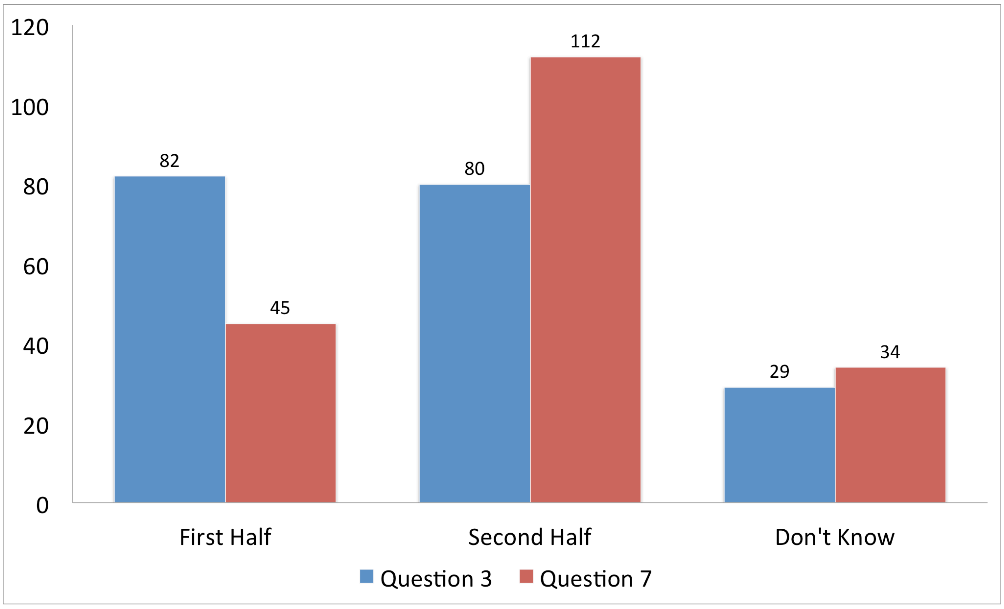


Figure 5.6. Data Breakdown for Questions 3 and 7

The data for questions 3 and 7 is presented in Figure 5.6. For question three the correct option was the first half of the year. The results show that listeners had difficulty answering question three with close to equal numbers of listeners selecting the first half of the year option and the second half of the year option. For questions seven a clear majority of listeners selected the correct answer, the second half of the year.

Two chi square goodness-of-fit tests were carried out on the data for question three. The first test included the results for the two halves of the year and the don't know option X^2 (2,191)=28.34, *p*<.001, V=.27. The second test considered only the results for the two halves of the year X^2 (1,167)=.02, *p*>.05, V=.01. The first test produced a significant *p* value of *p*<.001 and a medium effect size of V=.27. This suggests that the results are of high statistical significance and a medium level of practical significance. The second test produced a non-significant *p* value of *p*>.05 and a very small effect size of V=.01. This suggests that the number of listener's choosing the correct option, the first half of the year, over the incorrect option was statistically non-significant and of almost no practical significance.

Two chi square goodness-of-fit tests were carried out on the data for question seven. The first test included the results for the two halves of the year and the don't know option X^2 (1,191)=55.98, P<.001, V=.38. The second test considered only the results for the two halves of the year X^2 (1,157)=28.59, *p*<.001 V=.42. The first test produced a significant *p* value of *p*<.001 and a large[21] effect size of V=.38. This suggests that the results are of high statistical significance and a large level of practical significance. The second test produced a significant *p* value of *p*<.001 and a medium effect size of V=.42. This suggests that the number of listener's choosing the correct option, the second half of the year, over the incorrect option was of high statistical significance and medium practical significance.

The results for questions 3 and 7 suggest that a substantial number of listeners can accurately estimate what time-point they are hearing only 50% of the time (once in 2 questions). This results is no better than chance and suggests that listeners cannot effectively identify the half of the year in which a organised in terms of the temporo-spatial motion framework ends.

---

[21] Effect sizes for Cramer's V are interpreted on different scales determined by df. Hence .38 is a big effect size for 2 df but .42 is only a medium sized effect for 1 df in the results above.

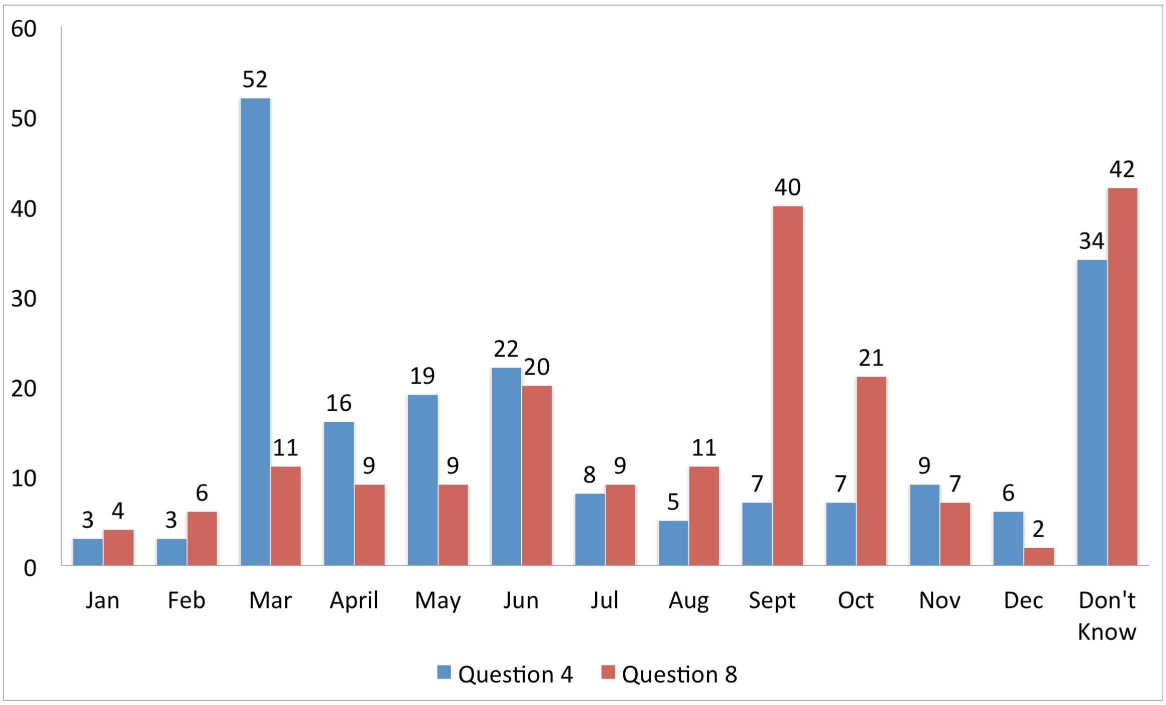


Figure 5.7. Data Breakdown for Questions 4 and 8

Questions 4 and 8 examined how precisely listeners could pin-point the exact months on which a sonification ended. The results are presented in Figure 5.7. 52 listeners pin-pointed the exact time point correctly for question 4 and 40 listeners pin-pointed the exact time point correctly for question 8. For question 8 the number of listeners selecting the correct option and the number of listeners selecting the don't know option were roughly equal.

Two chi square goodness-of-fit tests were carried out on the data for question four. The first test included the results for the twelve month options and the don't know option X^2 (12,191)=168.58, p<.001, V=.27. The second test considered only the results for the twelve month options X^2 (12,157)=158.43, p<.001, V=.3. The first test produced a significant p value of p<.001 and a medium effect size of V=.27. This suggests that the results are of high statistical significance and a medium level of practical significance. The second test produced a significant p value of p<.001 and a large effect size of V=.3. This suggests that the number of listeners selecting the correct option, March, over the eleven incorrect alternatives was of high statistical significance and large practical significance.

Two chi square goodness-of-fit tests were carried out on the data for question eight. The first test included the results for the twelve month options and the don't know option X^2 (12,191)=135.36, p<.001, V=.24. The second test considered only the results for the twelve month options X^2 (11,149)=95.1, $p<.001$ V=.24. The first test produced a significant p value of $p<.001$ and a medium effect size of V=.24. This suggests that the results are of high statistical significance and a medium level of practical significance. The second test produced a significant p value of $p<.001$ and a large effect size of V=.24. This further suggests that number of listeners selecting the correct option, September, were of high statistical significance and large practical significance.

The results for questions 4 and 8 indicate that a substantial number of listeners can accurately identify the month in which a sonification ends using the framework but an equally substantial numbers of listeners report that they did not know in which month the sonification ended. This indicates that while the framework is communicatively effective for some listeners, a similar number of listeners do not understand it.

5.13 How Far they are from the Beginning and End Points of a Sonification

For Q5 and Q9 listeners were presented with sonifications and asked to determine how many time-points had been presented since the beginning of the sonification. Listeners could choose any number between 1 and 12 and a 13th option of "Don't Know".

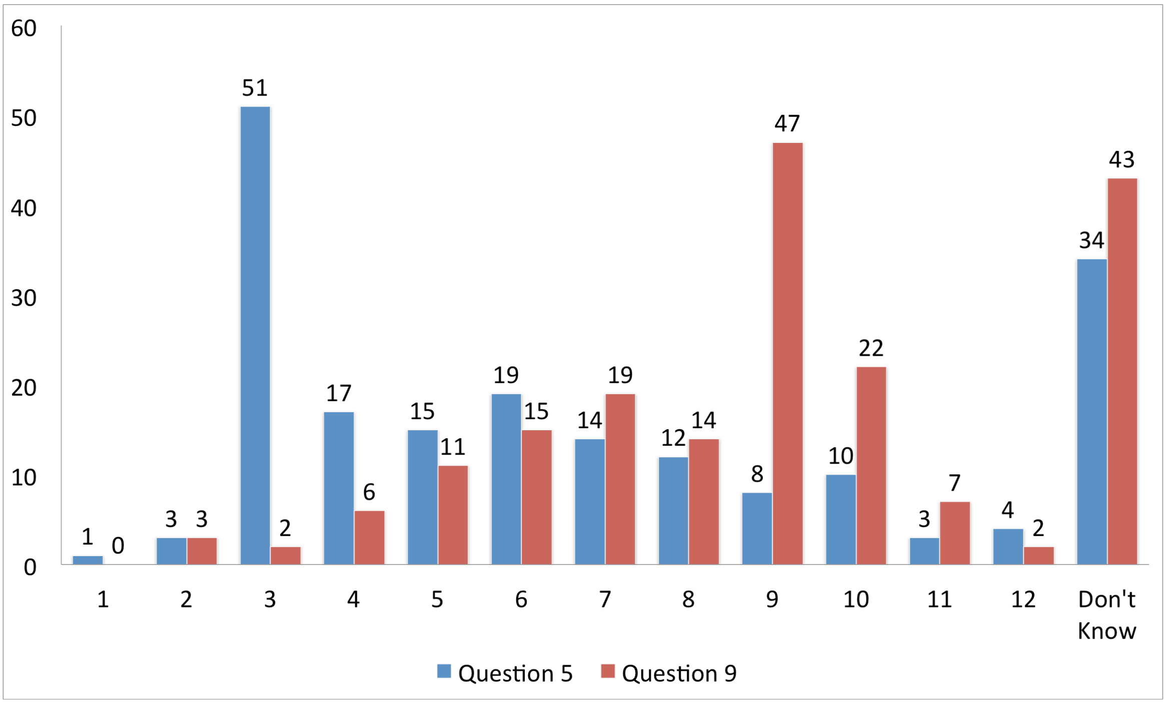


Figure 5.8. Data Breakdown for Questions 5 and 9

Questions 5 and 9 examined how precisely listeners could identify the number of months presented since the beginning of the sonification. The results are presented in

Figure 5.8. 51 listeners identified the number of months correctly for question 5 and 47 listeners identified the number of months correctly for question 9. For question 9 the number of listeners selecting the correct option and the number of listeners selecting the don't know option were close in value.

Two chi square goodness-of-fit tests were carried out on the data for question five. The first test included the results for the twelve number options and the don't know option X^2 (12,191)=160.95, $p<.001$, V=.27. The second test considered only the results for the twelve number options X^2 (11,157)=149.88, $p<.001$, V=.29. The first test produced a significant p value of $p<.001$ and a medium effect size of V=.27. This suggests that the results are of high statistical significance and a medium level of practical significance. The second test produced a significant p value of $p<.001$ and a large effect size of V=.29. This suggests that the number of listeners selecting the correct option, three, over the eleven incorrect alternatives was of high statistical significance and large practical significance.

Two chi square goodness-of-fit tests were carried out on the data for question nine. The first test included the results for the twelve number options and the don't know option X^2 (12,191)=186.54, $p<.001$, V=.29. The second test considered only the results for the twelve month options X^2 (11,148)=151.83, $p<.001$, V=.3. The first test produced a significant p value of $p<.001$ and a large effect size of V=.29. This suggests that the results are of high statistical significance and a roughly high level of practical significance. The second test produced a significant p value of $p<.001$ and a large effect size of V=.3. This further suggests that number of listeners selecting the correct option, nine, were of high statistical significance and large practical significance.

The results for questions 5 and 9 indicate that a substantial number of listeners can accurately estimate how many time points have been presented since the beginning of a sonification using the framework. However, a similarly substantial numbers of listeners report that they did not know how many time points had been presented in a sonification made using the framework. This indicates that while the framework is communicatively effective for some listeners, a similar number of listeners do not understand it.

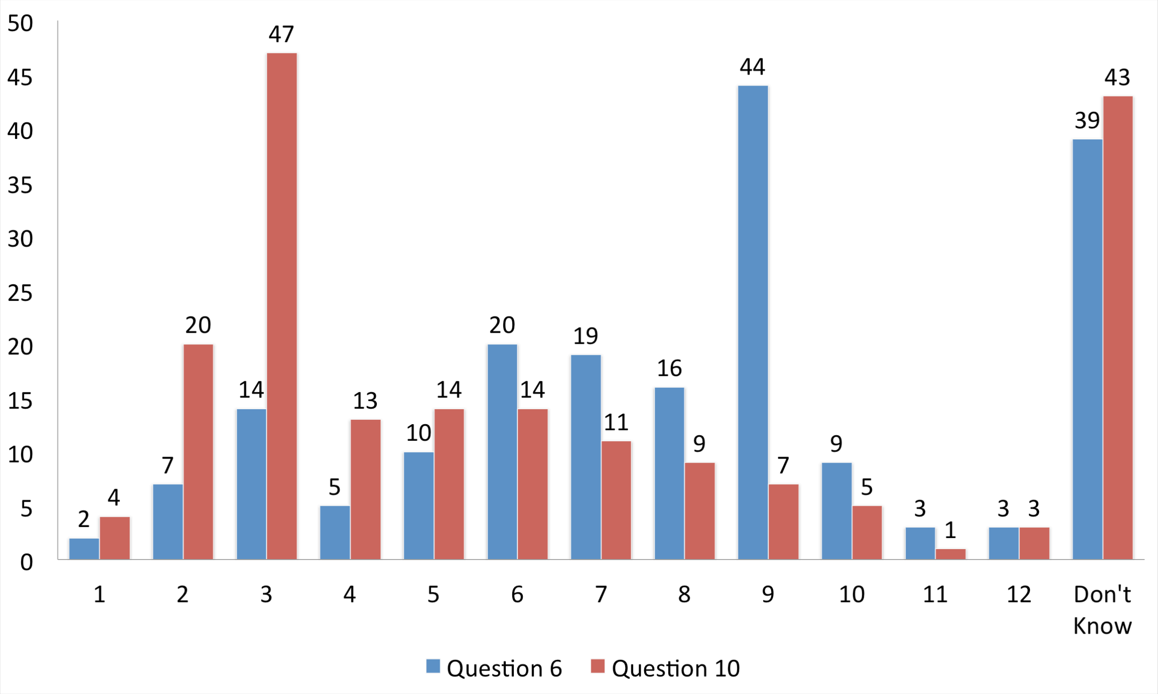


Figure 5.9. Data Breakdown for Questions 6 and 10

Questions 6 and 10 examined how precisely listeners could determine the number of time-points remaining in a sonification of time-series data. The results are presented in Figure 5.9. 44 listeners identified the correct number of remaining time points for question 6 and 47 listeners identified the correct number of time points remaining for question 10. For both question 6 and 10 the numbers of listeners selecting the correct options and the number of listeners selecting the don’t know options were similar in value.

Two chi square goodness-of-fit tests were carried out on the data for question six. The first test included the results for the twelve number options and the don’t know option X^2 (12,191)=145.7, p<.001, V=.25. The second test considered only the results for the twelve number options X^2 (11,152)=118.47, p<.001, V=.27. The first test produced a significant p value of p<.001 and a medium effect size of V=.25. This suggests that the results are of high statistical significance and a medium level of practical significance. The second test produced a significant p value of p<.001 and a medium effect size of V=.27. This further suggests that the number of listeners selecting the correct option, nine, over the eleven incorrect alternatives was of high statistical significance and a medium level practical significance.

Two chi square goodness-of-fit tests were carried out on the data for question ten. The first test included the results for the twelve number options and the don’t know option X^2(12,191)=168.58, p<.001, V=.27. The second test considered only the results for the twelve number options $X^2(11,148)=158.44$, $p<.001$, $V=.31$. The first test produced a significant *p* value of $p<.001$ and a medium effect size of $V=.27$. This suggests that the results are of high statistical significance and a medium level of practical significance. The second test produced a significant *p* value of $p<.001$ and a large effect size of $V=.31$. This further suggests that the number of listeners selecting the correct option, three, was of high statistical significance and large practical significance.

The results for questions 6 and 10 indicate that a substantial number of listeners can accurately determine the number of time-points remaining in a sonification using the framework. However, a similarly substantial numbers of listeners report that they did not know how many time-points remained in a sonification made using the framework. This indicates that while the framework is communicatively effective for some listeners, a similar number of listeners do not understand it.

5.14 How effectively can listeners judge the relationships between sequentially presented Data Points

For Q11 and Q12 listeners were presented with sonifications in which the rainfall data for two consecutive months was represented in sequence. They were then asked to determine which month had the highest data value. Listeners had 3 choices, the first month, the second month or "Don't Know".

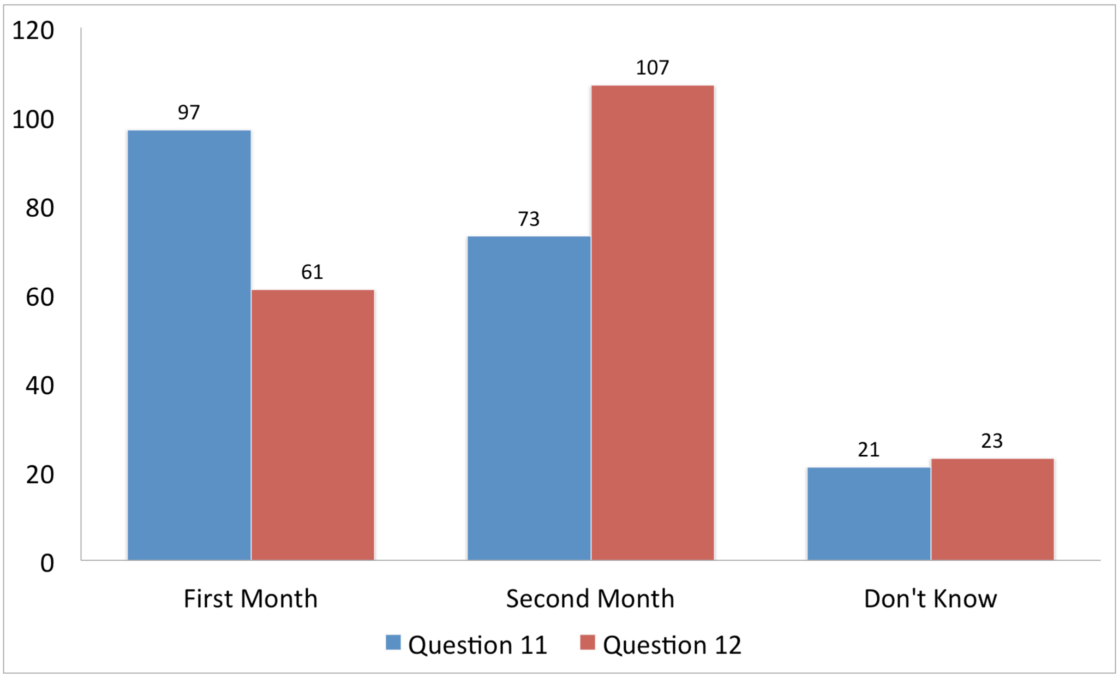


Figure 5.10. Data Breakdown for Questions 11 and 12

The results for questions 11 and 12 are presented in Figure 5.10. 97 listeners selected the correct option, the first month, for question 11. 107 listeners selected the correct option, the second month for question 12.

Two chi square goodness-of-fitness were carried out on the data for question eleven. The first test included the results for the two months the don't know option $X^2(2,191)=47.41$, $p<.001$, $V=.35$. The second test considered only the results for the two halves of the year X^2 $(1,170)=3.38$, $p>.05$, $V=.14$. The first test produced a significant *p* value of $p<.001$ and a large effect size of $V=.35$. This suggests that the results are of high statistical significance and a large level of practical significance. The second test produced a non-significant *p* value of $p>.05$ and a small effect size of $V=.14$. This further suggests that the number of listener's choosing the correct option, the first month, over the incorrect option was of low statistical significance and a small degree of practical significance. This suggests that listeners were confused by the sonification. The high level of statistical significance and large level of practical significance produced by the first test suggest that the number of listeners that did not know which option to choose were of high statistical significance and a large degree of practical significance.

Two chi square goodness-of-fit tests were carried out on the data for question twelve. The first test included the results for the two months and the don't know option X^2 $(2,191)=55.58$, $p<.001$, $V=.38$. The second test considered only the results for the two month options X^2 $(1,168)=.12.6$, $p<.001$, $V=.27$. The first test produced a significant *p* value of $p<.001$ and a large effect size of $V=.38$. This suggests that the results are of high statistical significance and a large degree of practical significance. The second test produced a significant *p* value of $p<.001$ and a small effect size of $V=.27$. This shows that the number of listener's choosing the correct option, the second month, over the incorrect option was of high statistical significance but showed only a small degree of practical significance.

These preliminary results suggest that 50% of the time (once in two questions), a significant number of listeners can effectively estimate the relationship between data-points that are sequentially presented using the temporo-spatial motion framework. This is not better than chance.

The results for questions 11 and 12 indicate that a substantial number of listeners can effectively estimate the relationship between data-points that are sequentially presented using the temporo-spatial motion framework 50% of the time (once in two questions). This result is no better than chance and suggests that listeners could not effectively understand the relationship between sequentially presented data-points sonified in terms of the temporo-spatial motion framework.

5.15 Discussion of Results.

Questions 1 and 2 tested the secondary aim B2 of the proposed framework, by exploring how effectively listeners can link data points to time-points within the proposed framework. For question 1 a statistically and practically significant number of did not know how data values and time-points were related in the stimuli. For question 2 a statistically and practically significant number of listeners misinterpreted the relationship between data values and time-points. These results provide strong evidence that the temporo-spatial motion framework should not be used to link data values to time-points. In the proposed TIME-SERIES DATA AS PHYSICAL MOTION METAPHOR, illustrated in Table 5.3, physical objects were mapped onto data points, changes in physical objects were mapped to changes in data value and vertical path of motion was mapped to value of the data. These results suggest that in the context of the temporo-spatial motion framework, these mappings do not reliably hold for the listener.

Questions 3, 4, 7 and 8 tested the primary aim A1 of the proposed framework by exploring how effectively listeners could determine what time-point they are hearing when data is sonified using the framework. Questions 3 and 7 tested whether or not listeners could identify the half of the year in which a sonification ended. The results suggest that listeners can accurately identify this only 50% of the time. This does not represent a reliable enough frequency. The results for question seven further suggest that in the cases where the number of listeners selecting the correct option was of high statistical significance and medium practical significance, a similar number of listeners still reported that they did not know how to use the framework. Questions 4 and 8 examined how well listeners could identify the exact months in which data points were presented. The results for these questions suggest that a number of listeners of high statistical significance and a medium degree of practical significance could use the framework to identify the time-point that they were hearing. However, a similarly large number of listeners reported that they did not know how to use the framework.

This suggests that the temporo-spatial motion framework may be useful in a limited manner for representing a time-series in sonification. These results suggest that the temporo-spatial motion framework is strongest when representing time-points and not as strong when expressing the relationships between data values.

In the original MOVING TIMES metaphor objects are mapped onto times, as illustrated in Table 5.1. In the MOVING OBSERVER metaphor locations on observer's path' are mapped onto times, as illustrated in Table 5.2. In the proposed TIME-SERIES DATA AS PHYSICAL MOTION METAPHOR, illustrated in Table 5.3, physical objects were mapped onto 'Data Points'. An analysis of the results for questions 1 and 2 indicates that this proposed mapping may be inaccurate. The analysis of questions 3,4,7 and 8 suggests physical objects might map to times, in the same way that they do for the MOVING TIMES and MOVING OBSERVER metaphor[22]. Further, this would suggest that temporary cessation of motion might map to pausing of time-series and speed of motion might map to speed of time-series presentation.

Questions 5, 6, 9 and 10 tested primary aim A2 of the proposed framework by exploring how effectively listeners could estimate their temporal distance from the beginning and end points of a time series. Questions 5 and 9 tested distance from the start point while 6 and 10 focused on distance from the end point. Listeners could determine both their distance from start and end points with a strong level of statistical significance and a medium to high degree of practical significance for each of the questions. However similar numbers of listeners also reported that they did not know how to determine their temporal proximity to the start and end points of a sonification using the framework. This suggests that the temporo-spatial motion framework might be used in a limited manner to give listeners a sense of temporal context in a sonification. The results for questions 9 and 10 are very similar, suggesting that once a listener has determined how many time-points they have already heard, their knowledge of the time-series, and/or the temporo-spatial motion framework, aids in accurately estimating the remaining time-points. Overall these results fall in line with what would be expected based on some of the metaphorical mappings suggested in the proposed TIME-SERIES DATA AS PHYSICAL MOTION METAPHOR and illustrated in Table 5.9.

The accuracy with which listeners judged the distance to sonification end points may suggest that they were performing some basic arithmetic using their prior knowledge of the data and their knowledge of the time-points to which they had just listened. This cannot be said with certainty. Nor can it be said with certainty that listeners were utilising

---

[22] Sonically there may be no difference between sonifications built around either of these two metaphors unless the designer chose to represent the sound of a listener in motion, e.g. walking sounds, or the sounds of vehicular travel, in a sonification built around the MOVING OBSERVER metaphor.

the knowledge that the passing data-objects with the closest proximity to them represented the mid-point of the data. Regardless it can be conclusively drawn that for this evaluation, listeners were able to determine their distance from the end-point of a sonification to a limited degree.

Source Domain (Physical Motion)		**Target Domain (Sonification)**
Movement past listener Front/Back	=	Passage of unit of time
Movement past listener Left/Right	=	Passage through data set
Listener's horizontal position	=	Midpoint of the data
Start/end of motion Left/Right	=	Beginning/end of time series
Start/end point of motion Front/Back	=	Beginning/end of time unit

Table 5.9. *Mappings supported by the findings for Question 3*

Questions 11 and 12 tested secondary aim B1 of the proposed framework by exploring how effectively listeners judge the relationships between data points presented sequentially within the temporo-spatial motion framework. The results suggest that listeners can accurately identify these 50% of the time. This does not represent a reliable enough frequency. These results further support the findings in questions 1 and 2, demonstrating that the temporo-spatial motion framework is not effective at representing relationships between data values. The results from evaluations 3 to 10 suggest that the framework is better suited to representing points in a time-series in a limited manner. This further indicates that a number of the metaphorical mappings suggested in the proposed TIME-SERIES AS MOTION metaphor, where physical objects map to data-points, changes in physical objects map to changes in data value, and vertical path of motion maps to value of the data, are incorrect.

The results suggest that physical objects map to times rather than data points and that changes in physical objects do not map to changes in data values. The proposed mapping from vertical height to data value is also invalidated by the evaluation. This further suggests that the temporo-spatial motion framework is more suited to representing the passage of points in a time series rather than the value of data points. It offers limited support to listeners in correctly identifying the time-points with which they are presented, and in estimating their temporal distance from the start and end points of a sonification. On the basis of these results a number of refinements are here presented in the TIME-SERIES DATA AS A PHYSICAL MOTION metaphor. Some of the results generated in this evaluation might be accounted for in terms of the limits of human memory. Echoic Memory is the form of sensory memory that stores sounds that a listener has just perceived. It operates by holding sounds in an unprocessed state until a following sound is presented in relation to which the original sound becomes meaningful (Clarke, 1987). This holding operation is estimated to have a capacity of roughly 4 seconds (Darwin *et al.*, 1972). Working memory is responsible for the short-term holding and processing of information and is often said to have a capacity of 7 ± 2 chunks, or discrete objects, of information (Miller, 1956) and is thought to decay after 10-20 seconds (Cowan, 1988; 1995). Each of the 12 individual stimuli presented in the sonification is of 9 seconds in length but the next stimuli in the sequence begins to sound 6 second into the prior stimulus. With these specifications framework can play back audio for 12 months in 1 minute and 15 seconds.

There are a number of factors surrounding the listener's abilities in terms of auditory memory, which are not addressed directly by the framework and may account for some of the difficulties listeners encountered during these evaluations. Echoic Memory is the form of sensory memory that stores sounds that a listener has just perceived. It operates by holding sounds in an unprocessed state until a following sound is presented in relation to which the original sound becomes meaningful (Clarke, 1987). This holding operation is estimated to have a capacity of roughly 4 seconds (Darwin *et al.*, 1972). Working memory is responsible for the short-term holding and processing of information and is often said to have a capacity of 7 ± 2 chunks, or discrete objects, of information (Miller, 1956)[23] and is thought to decay after 10-20 seconds (Cowan, 1988;1995). Each of the 12 individual stimuli presented in the sonification is of 9 seconds in length but the next stimuli in the sequence begins to sound 6 second into the prior stimulus. With these specifications framework can play back audio for 12 months in 1 minute and 15 seconds. The individual data-point stimuli are too long in duration to be held and processed by echoic memory and each stimulus roughly reaches the capacity of the lower limits of working memory. Similarly the entire sonification is too long to be held and processed as a whole in working. These violations of the limits of echoic and working memory may have contributed to the difficulties listeners encountered during the evaluation. Creating faster

---

[23] Miller points out that listeners can accurately make absolute judgement of the pitches of up to 6 tones but that musically trained listeners with perfect pitch might discern between 50 and 60. Listeners with such abilities may interpret the sonifications described in this chapter much differently.

variants of the framework might negate these effects. Such sonifications could present individual stimuli within a 4 second window in order to exploit the meaning-making capacities of echoic memory. They might also group portions of the time-series into a task specific number of chunks that fall within the 10-20 second window of working memory and observe the Miller limit. The way in which the limits of auditory memory are addressed may be different for different sonification designs depending on the task or usage for which the sonification is intended.

Source Domain (Physical Motion)		**Target Domain (Sonification)**
Physical object	=	Times
Movement past listener Front/Back	=	Passage of unit of time
Movement past listener Left/Right	=	Passage through data set
Listener's horizontal position	=	Midpoint of the data
Starting/ending point of motion Left/Right	=	Beginning/end of time series
Starting/ending point of motion Front/Back	=	Beginning/end of time unit
Temporary cessation of motion	=	Pausing of Time-Series
Speed of Motion	=	Speed of Time-Series presentation.

Table 5.10. *The Final Time-series Data as a Physical Motion Metaphor*

5.16 Refined Temporo-spatial Motion Framework specification

Having previously refined the TIME-SERIES DATA AS A PHYSICAL MOTION metaphor, the temporo-spatial motion framework is now refined in order to incorporate the results uncovered during empirical testing and better integrate with the embodied conceptual system implicit in meaning-making in sonification listening. The refined framework is solely concerned with representing the time-series data. This is achieved by denoting each individual time-point in the data with salient sonic objects of the designer's choice (e.g., pitch or timbre objects) to create sonic time-objects. Over the course of a single unit of time, a time-object will seem to pass by the listener, moving from a position in front of the listener to a position behind them. This trajectory, from front to back, represents the passage of a specific unit of time as illustrated in Figure 5.11. These passing time-objects are organised from left to right, relative to the listener's position, across the horizontal axis of the stereo spatial field. The time-object for the earliest point in the time-series appears to the extreme distant left of the listener. The mid point of the time-series appears at the listeners' location and the last point in the time series appears to the extreme distant right of the listener. This configuration, represented in Figure 5.12, creates an evolving matrix of data-driven temporal gestures. These gestures make one entire pass by the listeners' position along a y-axis (front to back) over the course of a single time-point and one entire pass by the listener's position along a z-axis (left to right) across the entire time series.

An example of the use of this framework to lend temporal context to a pitch mapped sonification of the rainfall data presented in Table 5.7 is included in the digital appendices for this chapter on the appendices DVD. The pitch-mapped sonification is presented to the centre of the stereo-spatial field while the framework is presented as described above.

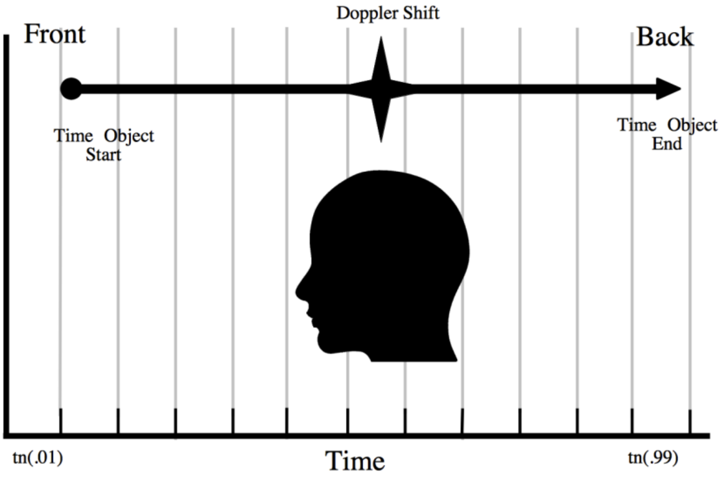


Figure 5.11. Refined Unit Passage from Front to Back

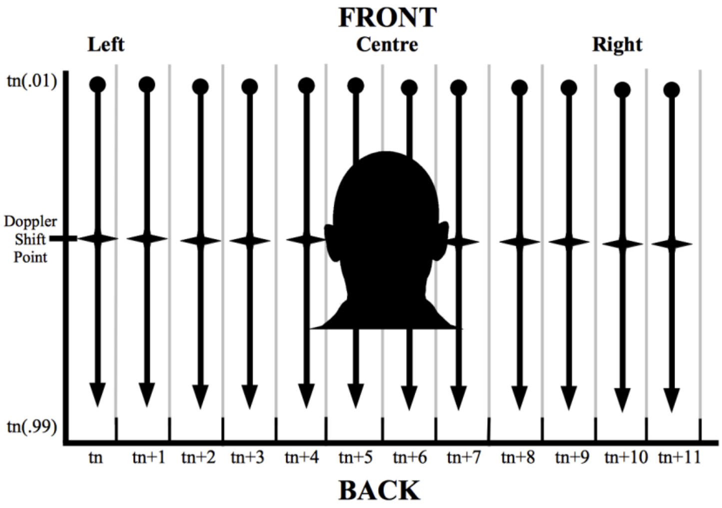


Figure 5.12. Refined Time-Series Passage from Left to Right

5.17 Limitations of the Framework

One limitation of the temporo-spatial motion framework is that it can only handle shorter time-series. The effectiveness of the temporo-spatial motion framework is constrained by two factors. The first is the resolution of the human auditory perceptual and cognitive system in localising sounds. The second is the designers' ability to represent a graded sense of distance to that auditory perceptual and cognitive system. The applications of the framework in this chapter focused on the rainfall data measured in mm at Dublin airport over the course of 2014 as represented in Table 5.7. There is a relatively short span of only 12 time-points represented in this data set. For data sets with a larger number of time-points listeners will not be able to discern small incremental changes in the spatial cues used. This can be addressed in a limited manner by treating ranges of sequential time-points as the sub-units of a finer timescale within the framework's time units. These sub-units are represented by tn(.01) to tn(.99) in Figure 5.6. These sub units can be configured to suit the designer's needs. This solution is still limited by the resolution of the perceptual and cognitive auditory system. It can only be effective up to the threshold of minimum audible angle where listeners can still distinguish one time-point from the previous time-point. As such the framework is better suited to providing a topological gist of the time-series to a listener.

5.18 Conclusions

This chapter has introduced the TIME-SERIES DATA AS SPATIAL MOTION metaphor and the temporo-spatial motion framework, in an attempt to offer a solution to the problem of temporal context in time-series sonification. Sonification stimuli in the format prescribed by the proposed framework were submitted to empirical evaluation. The results of this evaluation were analysed and both the framework and the metaphor were further developed and refined to reflect these results. The results indicate that the temporo-spatial motion framework offers the listener a limited sense of temporal context during sonifications of time-series data by aiding the listener in estimating which time-point they are currently hearing, how many time-points they have already heard, and how many time-points they have left to hear.

Chapter 6: Summary and Discussion

6.1 Introduction

This chapter presents a discussion of the work, findings and conclusions presented throughout this thesis. It opens with a set of discussions that summarise the implications and conclusions of each chapter before presenting further practical and theoretical conclusions which are drawn from this thesis as a whole. The chapter revisits the original thesis statement before closing with a discussion of the conclusions and the value of their contribution to the field of auditory display.

6.2 Chapter 1: Embodied Sonification

This chapter lays out a general research direction and establishes the theoretical underpinnings of this thesis. It introduces the motivation behind the work, and offers the following thesis statement: ‘An embodied cognitive approach to meaning-making and aesthetic practice can support the design of communicatively effective sonifications’. Chapter 1 discusses embodied aesthetics and the promise that an embodied approach to meaning-making in sonification may hold, before presenting a definition which frames sonification as a meaning-making activity. What are here termed the representational and embodied models of meaning-making are then examined and discussed before a consideration of some of the embodied meaning-making faculties modelled by embodied cognitive science is undertaken. The historical development and conceptual underpinnings of both models are discussed, and it is argued that the computationalist model of meaning-making is insufficient to account for the embodied meaning which underlies the aesthetic dimensions of sonic experience. An embodied approach to meaning-making in sonification that can harness these aesthetic dimensions is proposed instead.

A number of popular sonification techniques are also considered in relation to those meaning-making models. Parameter Mapping Sonification (PMSon), the use of environmental sound, and the use of musical structures are highlighted as sonification techniques that might be used to exploit embodied meaning-making. PMSon is applied as a method in Chapter 3, musical structures are used in Chapters 3 and 4 and environmental sounds are used in Chapter 4.

Scientifically and aesthetically oriented approaches to sonification are also investigated in terms of both models and it is suggested that, in general, scientifically oriented approaches acknowledge embodied meaning in sonification but tend to frame the listener's meaning-making process as an information processing task. Typically, it is proposed that aesthetics oriented approaches can more effectively account for an embodied model of meaning-making in sonification.

The chapter closes with a discussion which further highlights some of the limitations of applying an exclusively computationalist paradigm to sonification, e.g., increased cognitive load, and an inability to account for the aesthetic dimensions of sound.

This chapter makes a number of contributions to sonification research and practice. It makes the argument that an embodied approach to meaning-making can provide a context in which to address the demand, identified by Degara *et al.* (2013) and Hermann (2008), for an empirically responsible and testable framework for sonification and the demand, identified by Barrass (2012), Barrass and Vickers (2011), Serafin *et al.* (2011), Vickers and Hogg (2006), Vickers (2005), to harness the aesthetic dimensions of sound to create more effective sonifications. It also offers a critique of the limits of the computationalist mode of meaning-making in the field, and introduces a novel embodied model to account for these shortcomings. This is critical to the development of an understanding of the cognitive factors at play during sonification listening, as called for by (Vickers and Hogg, 2006; Neuhoff, 2011; Gossman, 2010; Worrall, 2009; Neuhoff and Heller, 2005; Walker and Kramer, 2004) and to the development of a coherent and unified theory of sonification as called for by (Walker and Nees, 2011).

6.3 Chapter 2: The Embodied Sonification Listening Model

Chapter 2 examines the embodied cognitive underpinnings of sonification listening. This chapter presents the Embodied Sonification Listening Model (ESLM) as the first dedicated theoretical model of embodied meaning-making in sonification listening. The model is also flexible enough to account for the computationalist model of meaning-making. The ESLM introduces two new conceptual schemes for thinking about sound in a sonification context: embodied sonic dimensions and embodied sonic complexes. The ESLM suggests that during sonification listening a listener will associate the embodied sonic complex with the phenomenon recorded in the data and interpret changes along the embodied sonic dimensions as changes along the measured dimensions of the original data. As such, listeners experience a sonification as a sonic metaphor for the data. How the data is mapped to sound depends on the listeners' background context of embodied schematic knowledge.

The model can serve as a basis for deriving empirically testable hypotheses about meaning-making in sonification listening. Chapter 2 presents strong empirical evidence in support of a number of hypotheses about the cognitive strategies listeners employ during sonification listening. Evaluations of these hypotheses showed that listeners' understanding of the data determines how they experience a sonification. When listeners believe they are listening to number data, they tend towards mapping increases in tempo and pitch to increases in the data and *vice-versa*. When they believe that the sonification represents attribute data they map decreases in tempo and pitch to increases in the data, and vice-versa, with a higher than expected frequency. This supports the assertion that listeners draw on their previous embodied schematic knowledge of the data dimensions in question to make sense of a sonification. It further shows that listeners map data-dimensions to embodied sonic dimensions during sonification listening. This is a noteworthy finding that provides support for the Embodied Sonification Listening Model (ESLM) which predicts this behaviour. These findings also begin the process of describing how different categories of data are naturally interpreted by a listener without any training or instruction, during sonification listening. It shows that increases in amount and energy measurements map directly to increases along pitch and tempo dimensions while increases in physical, spatial, and weight measurements map to decreases along pitch and tempo dimensions. This information is of importance to the field as it offers guidance on how to create mapping strategies that best fit with the listener's innate cognitive meaning-making faculties allowing the designer to leverage these abilities rather than requiring listeners to learn an abstract set of rules for interpreting a sonification. As such, it holds a number of important implications for the practice of sonification. It implies that effectively communicative data-sonifications can be achieved by designing sonification solutions that are suited to the listener's embodied meaning-making faculties as described in the ESLM. The remainder of the thesis is spent exploring this premise and providing support for the ESLM.

6.4 Chapter 3: Embodied Sonic Dimensions - Vocal Gestures

Chapter 3 provides an exploration of the embodied sonic dimensions introduced by the ESLM. The chapter opens with a discussion of PMSon and the mapping problem. It then suggests that embodied sonic complexes and embodied sonic dimensions may provide more useful conceptual measures for sonification than traditional acoustic and psychoacoustic dimensions. The chapter selects vocal gestures as a domain of embodied sonic complexes and explores the embodied sonic dimensions that vocal gestures afford a listener. Vocal gestures and stereo space are defined and explained, before descriptions of two prototyping platforms used in the development of hypotheses related to the embodied sonic dimensions afforded by vocal gestures are presented. A piece of data-driven music titled *The Human Cost* is also presented in this chapter, the compositional process of which allowed for the further development of hypotheses about the embodied sonic dimensions of vocal gestures. Four hypotheses are put forward on the basis of this exploratory research. Each of these is then tested empirically through listener evaluations. The results of these evaluations strongly imply that embodied sonic dimensions can be used to effectively communicate data to a listener. The parameters required to model each dimension are described and sonifications which are written in Csound and map data to these dimensions are supplied. The seven dimensions explored in the chapter are broken into three categories - embodied schematic, vowel based, and emotional. The dimensions are Speed, Size, Amount, Vowel Attribute, Emotion, Tension and Texture.

Chapter 3 makes a number of significant contributions to the field of auditory display. It provides empirical support for the validity of the ESLM and the use of embodied sonic dimensions and embodied sonic complexes as conceptual tools for extending the embodied meaning-making paradigm to sonification. This chapter also outlines and empirically evaluates the communicative effectiveness of a set of embodied sonic dimensions. The testing undertaken, and the sonifications and data-driven musical compositions created, highlight the usefulness of the ESLM as a guiding force for sonification design and the potential of the embodied sonic dimension for effectively communicative sonification.

Chapter 3 also proposes that the mapping problem can be avoided by mapping independent data streams to independent embodied sonic complexes. This is a significant contribution towards solving a problem which has been identified as a serious obstacle to the effective application of PMSon, and as such the development of the field as a whole (Worrall, 2013).

Chapter 3 examines the interplay between aesthetic practice, embodied cognition and sonification. It argues that both sonification cognition and aesthetic cognition are concerned with meaning-making and are driven by the same cognitive meaning-making faculties. The data-driven composition *The Human Cost*, derived from a number of socioeconomic data sets, lends further support to this claim as it makes use of the listener's capacity for aesthetic experience to communicate data to them. This piece was chosen for performance at Ireland's National Concert Hall by the Contemporary Music Centre .

Chapter 3 highlights the effectiveness of the human voice, and vocal gestures in particular, as a domain of embodied sonic complexes which contain effectively communicative embodied sonic dimensions for sonification. This is especially true in the sonification of socio-economic data, where the voice presents an appropriate metaphor for the people measured in the data in both a figurative sense and in the more literal sense described by the ESLM.

Chapter 3 demonstrates a large number of embodied sonic dimensions inherent in vocal gestures, which can be used for effectively communicative sonification. It also describes the empirical research methods required to uncover these dimensions.

6.5 Chapter 4: Embodied Sonic Complexes – Embodied Soundscape Sonification Framework

Chapter 4 explores the concept of the sonic complex introduced by the ESLM. The sonic complexes used in this chapter are derived from environmental soundscapes. The chapter presents the embodied soundscape sonification framework as a solution to the question of how best to utilise the communicative capacity of the soundscape for sonification. The chapter opens with a presentation of motivations followed by a discussion of the embodied cognitive underpinnings of the soundscape and the role of the soundscape in sonification. The concept of an embodied soundscape sonification is then introduced as an approach to sonification that unites both the ecological and embodied paradigms through the intermediary of the ESLM. A number of prototyping platforms that were used for exploratory research are introduced. This is followed by a discussion of *Idle Hands*; a data-driven soundscape, the composition which allowed for a further exploration of the methods by which data can be mapped to soundscape to exploit embodied auditory cognition. Based on findings uncovered during explorations undertaken with the prototyping platforms, four candidate frameworks for embodied soundscape sonification are then presented. Each framework presents an approach to dealing with embodied sonic complexes and embodied sonic dimensions. Strategies for synthesising these frameworks are then discussed and a demonstration of one such solution is presented. This is followed by empirical listener evaluations of all four candidate frameworks to determine which is the most communicatively effective. The results of this evaluation are discussed before the chosen embodied soundscape sonification framework is put forward alongside a consideration of its limitations. The chapter finishes with a discussion and set of concluding remarks.

This chapter makes a number of practical contributions to sonification practice and research. It puts forward the embodied soundscape sonification framework. This is an empirically supported framework for soundscape sonification that exploits the embodied nature of meaning-making during sonification listening as described by the ESLM. It also presents empirical evidence which suggests that, in certain contexts, the framework can be more effective at communicating data to a listener than pitch-mapped sine tone sonifications. This is important when the prevalence of pitch mapping across sonification research and practice is considered (see Grond and Berger, 2011).

This framework treats the embodied sonic complex as a metaphor for the data source, as described in the ESLM. Three alternate frameworks, two of which did not treat sonic complex as a metaphor for the data source, were submitted to empirical listener evaluations and found to be less effective than the proposed soundscape sonification framework. As such this finding offers support for the ESLM.

The chapter further highlights how an embodied approach to sonification can inform aesthetic practice, in the context of data-driven music through the two data-driven compositions presented within - *Idle Hands* and *Doom & Gloom*, both of which have been chosen for performance in a concert hall setting, with the latter being performed extensively throughout 2015 including performances at Ireland’s National Concert Hall, and during the 2015 sound-night for the International Conference for Auditory Display in Graz, Austria.

6.6 Chapter 5: Embodied Sonic Metaphor - The Temporo-spatial Motion Framework

Chapter 5 explores the role of the sonic metaphor in sonification as described in the ESLM. The chapter provides a solution to the problem of representing temporal context in the sonification of time series data, by imposing a formal organisation upon sonification that is informed by the embodied cognition literature, and is congruent with the ESLM. The chapter opens with a discussion of motivations, and the embodied cognitive underpinnings of time and music. A hypothetical system of conceptual metaphorical mappings that conceptualises the time-series in a data set in terms of motion, and might underpin sonification listening, is then proposed. From here the aims of the temporo-spatial motion framework are presented and discussed and the proposed framework itself is defined. Strategies for synthesising the framework, using spatial audio techniques like reverb, panning and Doppler processing are explored. Also, an application of the framework to time-series rainfall data is presented, whereby rainfall data recorded at Dublin Airport during 2014 is mapped to pitch and is spatially organised on the basis of the framework. This is followed with an experimental listener evaluation and consideration of results that serves to determine which aspects of the proposed framework are communicatively effective. These results suggested that the framework allowed listeners to gauge their position within a time-series with a high level of accuracy and that data points and time-points should be presented in isolation. The refined temporo-spatial motion framework is then presented at the end of this chapter before a discussion of its limitations and some closing conclusions.

This chapter puts forward the first empirically supported conceptual metaphor proposed to underlie the sonification of time-series data. The temporo-spatial motion framework is built around this conceptual metaphor and submitted to empirical listener evaluation. The results of this evaluation are used to refine both the framework and the original Time-series as Motion Metaphor. The Time-Series as Motion Metaphor and the temporo-spatial motion framework represent a contribution to sonification research and practice that is of practical use. The metaphor describes how listeners draw upon their understanding of motion, at both conscious and sub-conscious levels, to understand and reason about time-series data. The framework offers an empirically verified technique for harnessing the potential of that conceptual metaphor for the effective sonic communication of data to a listener. Both the metaphor and the framework could be of use to researchers and practitioners working with time-series data by allowing them to sonify their data in a format that the listener's embodied meaning-making faculties can intuitively understand. The chapter provides further support for the effectiveness of the ESLM as a guide for sonification design by illustrating how the sonic metaphor described by the ESLM can be exploited for the effective communication of time-series data to a listener.

6.7 Further Practical and Theoretical Contributions

This thesis opened with a discussion of some of the challenges facing sonification research and practice. In summary, these were the need for more scientifically rigorous approaches in the field (Hermann, 2008; Degara *et al.,* 2013), a call to exploit the aesthetic dimensions of sound in greater depth (Barrass, 2012; Barrass and Vickers, 2011;

Serafin *et al.*, 2011; Vickers and Hogg, 2006; Vickers, 2005), an understanding of cognition in sonification (Vickers, 2013; Neuhoff, 2011; Gossman, 2010; Worrall, 2009; Neuhoff and Heller, 2005; Walker and Kramer, 2004), and a need for theoretical tools specific to sonification (Walker and Nees, 2011). Each of these challenges were addressed in this thesis. The Embodied Sonification Listening Model provides a description of the cognitive faculties involved in meaning-making during sonification listening. The ESLM, Embodied Sonic Dimensions, and Embodied Sonic Complexes are conceptual tools which also contribute towards the need for a theoretical framework specific to sonification. They provide a basis for exploring the aesthetic dimensions of sound in an empirically rigorous manner through the development of empirically testable hypotheses *via* exploration with prototyping platforms and the compositional process. Evidence for the effectiveness of the ESLM, Embodied Sonic Dimensions and Embodied Sonic Complexes as tools for designing communicative sonifications is repeatedly presented throughout this thesis in the results of the empirical listener evaluations.

This thesis makes significant contribution to sonification research and practice. It provides an in-depth analysis of the computationalist approach to meaning-making and the limitations it imposes upon sonification. It also presents new embodied definitions of sonification and sonic meaning-making which frames sonification listening as a meaning-making activity and meaning-making as an embodied activity. This is also a useful practical contribution because it highlights an alternative paradigm for both sonification research and contemporary research in sound and music, where the role of body in cognition is taken seriously, and so meaning-making is more fully accounted for.

The mapping of data to embodied sonic complexes and dimensions provides a practical and easily implemented solution to the mapping problem. The use of data to control the embodied schematic relationship between multiple sonic complexes, as explored in the embodied soundscape sonification framework, provides an elegant solution that utilises the communicative effectiveness of environmental soundscapes for sonification. The temporo-spatial motion framework exploits the listeners' conceptual metaphorical mappings to express time-series data in a direct and intuitive manner. Each of these represent a significant contribution to sonification practice.

Another important outcome of the approach explored in this thesis is the intuitive nature of the sonifications created. Approaches to sonification based on the computationalist model of meaning-making often require that users learn the meanings of different sonic tokens presented in a sonification. This is because such sonifications present a listener with a set of sonic symbols that are modulated on the basis of data changes. The intended meanings of these symbols then must be learned in order for a listener to understand the sonification. The embodied approach to sonification aims to ground the sonic symbols presented to a listener in a sonification or auditory display context. This approach exploits the listener's innate embodied cognitive meaning-making faculties to express data in a more direct manner. These sonifications require minimal training and learning on the part of a listener. This is a contribution of practical use to the design of sonifications for a general audience of listeners. Minimising the amount of training required for the listener to interpret a sonification is a critical step along the road to making sonification an accessible and widely popular medium for understanding the rich data flows of increasingly networked societies.

During the exploration stages of this project, described in Chapters 3 and 4, a number of data-driven musical compositions were devised and realised. The aim was to apply embodied cognition principles and sonification techniques to the composition of data-driven music and to use the compositional process to explore the communicative effectiveness of some of the aesthetic sonic dimensions made available by an embodied approach to sonification. These compositions were invaluable in guiding the empirical investigations that would follow. *Doom & Gloom* and *The Human Cost* were both performed at Ireland's National Concert Hall as part of the Contemporary Music Centre's Spring 2015 Salon Series; the first salon of its kind to feature electronic music solely. It was also performed at the 2015 Sound and Music Computing conference at the National University of Ireland Maynooth, the 2015 International Conference on Auditory Display at the Kunst Uni in Graz, Austria and the 2015 Irish Sound Science and Technology Convocation at Dance Limerick, Ireland. *Idle Hands* was performed at the 2014 Irish Sound Science and Technology Convocation at the National University of Ireland Maynooth. All three compositions were also performed in an impromptu fashion at Discover Research Dublin 2015 at Trinity College Dublin. These pieces illustrate the ways in which an embodied approach to sonification can inform aesthetic practice in the context of data-driven music.

The research described in this thesis has produced and motivated a number of academic papers and presentations. Embodied Aesthetics in Auditory Display was a journal article authored by Stephen Roddy and Dermot Furlong. It was published in Organised Sound on February 27, 2014 (see Roddy and Furlong, 2014). This article discussed how an aesthetic framework for auditory display design might solve the mapping problem. The work described in this article was further developed and refined and would go on to motivate and guide the research presented in Chapter 3 of this thesis. Embodied Cognition in Auditory Display was a paper for oral presentation authored by Stephen Roddy and Dermot Furlong. It was presented at the 19th International Conference on Auditory Display, Lodz, Poland July 9th 2013 (see Roddy and Furlong 2013a). The paper discussed embodied cognition and its role in sonification before discussing how Johnson's twin-pan balance schema could be used to lend an intelligible structure to a sonification. Part of this paper would go on to motivate some of the theory discussed in Chapter 1 and it would also go on to motivate the work presented in Chapter 4. Sonification Listening an Embodied Approach was a paper for poster presentation authored by Stephen Roddy and Dermot Furlong and presented at the 21st International Conference on Auditory Display, Graz, Austria July 9th 2015 (see Roddy and Furlong, 2015b). It presented an alternate version of the sonification listening model, and described the experiment undertaken in Chapter 2 of this thesis. This model was further refined after the authoring of the paper and is also presented in Chapter 2 of this thesis. Embodied Affordances in Auditory Display was a paper for poster presentation authored by Stephen Roddy and Dermot Furlong. It was presented at the 12th Sound and Music Computing Conference, Maynooth, Ireland July 31st, 2015 (See Roddy and Furlong, 2015a). The paper extends Gibson's theory of affordances to account for the embodied nature of auditory cognition and applies the theory to auditory display. The size experiment contained in Chapter 3 of this thesis is then presented and discussed in the context of this framework. Rethinking the Transmission Medium in Live Computer Music Performance was a paper for oral presentation authored by Stephen Roddy and Dermot Furlong (see Roddy and Furlong 2013b). It was presented at the Irish Sound Science and Technology Association August 20, 2013. The paper discussed computationalist approaches to music making in a live computer music context contrasting the disconnect between performer and audience with the mapping problem in sonification. It then discussed a number of early exploratory investigations intended to explore how embodied cognition design principles might inform mapping strategies in the context of sonification and live computer music. Some of the theoretical background in this paper motivated and guided the work presented in Chapters 1 and 3. The exploratory techniques employed here would also be adopted as a research approach for Chapters 3 and 4. Sonification and the Digital Divide was a presentation authored by Stephen Roddy and presented at the Digital Material Conference, Galway, Ireland, May 21st 2015. The presentation explored the history of sound and its ontological status in Western culture before reviewing sonification and contextualising the author's work in relation to the field. Data Listening was a poster presentation authored by Stephen Roddy and presented at Discover Research Dublin on September 25th, 2015. It presented and discussed the results for both the experiment presented in Chapter 2 and the size experiment presented in Chapter 3 contextualising these within the wider frame of the work presented and discussed in this thesis.

6.8 Revisiting the Original Thesis Statement.

The original thesis statement for this project was presented as follows:

An embodied cognitive approach to meaning-making and aesthetic practice can support the design of communicatively effective sonifications.

The ESLM provided this thesis with a model of sonification listening from the embodied cognitive point of view. This acted as an aid to the research and creation of communicatively effective sonifications that could exploit the embodied nature of auditory cognition. From there, the thesis was able to offer some empirically supported, and practically significant, observations on the role of embodied knowledge, the knowledge a subject derives from their everyday embodied experiences, in the interpretation of pitch and tempo dimensions in sonification listening. The thesis then explored and evaluated strategies for mapping data to the communicative dimensions of synthesised vocal gestures. It then employed empirical evaluations of potential frameworks to create the embodied soundscape sonification framework. This framework organises soundscape sonification along embodied dimensions. It also presented the Temporo-spatial Motion framework, which aimed to exploit the listener's conceptual metaphorical mappings, from the time domain to the spatial motion domain, in order to give listeners a sense of temporal context during signification listening.

This thesis also highlights how an embodied approach to sonification can inform aesthetic practice, in the context of data-driven music. The three data-driven musical pieces presented in this thesis paired compositional choices informed by the embodied cognition literature with sonification techniques to produce musical results.

A number of novel approaches to sonification research, exploration with prototyping platforms, research via composition and the empirical evaluation of hypothesised conceptual metaphors were also explored. This thesis has successfully and exhaustively explored its original thesis statement and made a large and important contribution to the field of sonification in the process.

6.9 Limitations of the Work

This thesis is focused on exploiting the listener's embodied meaning-making faculties for sonification design and does not explore other aspects of embodiment and embodied cognition that might be of relevance to sonification and auditory display. The empirical evaluations undertaken in this thesis were performed online and as such results are of most significance in the context of the relatively large international audiences of non-specific online sonification listeners listening across a wide variety of devices and of more limited generalisability beyond that.

Chapter 1 focuses on reviewing sonification research that can be clearly defined as either aesthetic or artistic in nature and sonification work that can be defined as scientific in nature. This meant that any sonification research that fell outside of these definitions was not included in the discussion thus limiting the scope of the chapter. Chapter 2 explores only a limited number of attribute schemata and data-types respectively. It also focuses exclusively on pitch-mapped sonifications. Therefore the results generated are limited in their generalisability. Chapter 2 also introduces the concept of the embodied sonic complex and the embodied sonic dimension but only two such constructs are explored throughout the remainder of the thesis. Chapter 3 focuses exclusively on vocal gestures and on a relatively small number of embodied sonic dimensions contained therein. The results produced by Chapter 3 therefore are unique to the domain of vocal gestures as set out in the chapter. Chapter 4 focuses exclusively on the environmental soundscapes as sources of embodied sonic complexes. It considers only the balance and twin-pan balance schemata as organisational strategies for embodied soundscape sonification and only two methods, the conceptual metaphor and conceptual blend, or exploiting the embodied dimensionality of sound. There is a wealth of embodied schemata, conceptual metaphors and conceptual blends that could be drawn on to inform the embodied soundscape sonification framework. Furthermore, the soundscapes created in Chapter 4 were composed from materials derived from a large range of unique source environments. These were then reassembled to create new soundscapes that did not accurately reflect any specific real-world location. The temporo-spatial motion framework presented in Chapter 5 is best suited to communicating a sense of topological gist of the time-series to a listener and is of limited use beyond that context.

6.10 Future Work

Further research is required to better understand the embodied components of sonification listening and to further refine the ESLM. For example, the ESLM describes sonification listening in terms of conceptual metaphorical mapping at the cognitive level. The evaluations for Chapter 2 explored listeners' interpretations of twenty four data types and used stimuli which changed in either pitch or tempo. Future research might consider more data-types and investigate how listeners perceive changes in more complex dimensions such as the embodied sonic dimensions explored in Chapter 3 and 4. The results for Chapter 2 suggested and that negative polarity mappings were more likely for the data types listed in Table 2.4 and most likely for depth. It was theorised that this might be due to listeners conceptualising depth in terms of negative vertical motion. Further research might investigate whether this is in fact the case and also determine what other data types are interpreted in a similar manner.

Results obtained in the Chapter 3 evaluations also suggest further research. While Chapter 3 explored a number of embodied dimensions specific to vocal gestures, there are many more dimensions that are not explored in this thesis. There are also many more embodied sonic complexes, sounds that present communicative embodied sonic dimensions to a listener. Future research might focus on creating a taxonomy of embodied sonic complexes which details the embodied dimensions those sounds offer to a listener. The evaluations suggested that the synthesis parameters for controlling the perceived size, speed, texture, tension and spatial orientation of a vocal gesture are effective to a limited extent but further research is needed to determine excess synthesis parameters that could be used to better control these dimensions. Future research is also needed to determine how emotion and the perception of proximity can be effectively controlled in a vocal gesture and also to further exploit the Doppler effect as a means of demarking space in a sonification context, as the results for the Doppler test were promising.

As discussed in section 6.9 the embodied soundscape framework discussed in Chapter 4 might be further expanded by accounting for a larger range of embodied schematic organisational strategies and new methods for exploiting the embodied dimensionality of sound. Further research is required to achieve this. Results obtained in Chapter 4 suggest that design choices not accounted for by the embodied soundscape sonification framework can impact the communicative effectiveness of the sonifications produced to the specifications of the framework. Further research is required to develop the design practices that might determine and account for these factors.

Further research is also required to establish whether the TIME-SERIES DATA AS PHYSICAL MOTION proposed in Chapter 5 is empirically supported and to determine how well the temporo-spatial motion framework works in the context of specific sonification tasks and to refine the framework accordingly.

The data-driven compositions presented in this thesis were used to help guide the development of hypotheses about how embodied dimensions of sound can be exploited in a sonification context. As such they were not empirically evaluated to determine their own capacity to communicate information about a data source or to determine their aesthetic merit. Further research is needed to examine the link between aesthetic merit and communicative effectiveness in the context of data-driven music.

The TIME-SERIES DATA AS PHYSICAL MOTION metaphor proposed in Chapter 5 requires further research to determine whether or not it is empirically supported. Further work is also required to determine how well the temporo-spatial motion framework works in the context of specific sonification tasks and to refine the framework accordingly.

6.11 Conclusion

This chapter summarises the research findings and contributions made by each chapter of this thesis, before reviewing some general practical and theoretical ramifications for sonification research and practice, and revisiting the thesis statement offered at the beginning of this work.

References

Alba-Ferrara L., Hausmann M., Mitchell R.L. & Weis S. (2011). The neural correlates of emotional prosody comprehension: disentangling simple from complex emotion. *PLoS ONE 6*(12).

Antle, A.N., Corness, G. & Bevans, A. (2011). Springboard: designing embodied schema based embodied interaction for an abstract domain. In England, D. (Ed.) *Human-Computer Interaction Series: Whole Body Interaction* (pp 7-18*)*. Berlin, Heidelberg: Springer.

Armstrong, D. F. (2002). *Original signs: gesture, sign, and the sources of language.* Washington D.C.: Gallaudet University Press.

Armstrong, D. F., Stokoe, W. C., & Wilcox, S. E. (1995). *Gesture and the nature of language.* Cambridge: Cambridge University Press.

Baier, G. & Hermann, T., (2004). The sonification of rhythms in human electroencephalogram. In: *Proceedings of the 10th International Conference on Auditory Display (ICAD 2004). Sydney, Australia, July 6-9, 2004.*

Baier G, Hermann T, & Stephani U. (2007). Multi-channel sonification of human EEG. In: *Proceedings of the 13th International Conference on Auditory Display (ICAD 2007), (pp. 491–496), Montréal, Canada, June 26-29, 2007.*

Baker, G. & Morris K.J. (1996). *Descartes' Dualism.* London: Routledge.

Barrass, S. (1997). *Auditory information design.* (Doctoral dissertation) Australia National University, Australia.

Barrass, S. (2012). The aesthetic turn in sonification towards a social and cultural medium. *AI & Society*, *27*(2), 177-181.

Barrass, S. & Vickers, P. (2011). Sonification design and aesthetics. In Hermann, T., Hunt, A., & Neuhoff, J. G., (Ed.) *The Sonification Handbook*, (pp145–171). Berlin, Germany; Logos Publishing House.

Barrett, N., & Mair, K. (2014). Sonification for geoscience: listening to faults from the inside. In *EGU General Assembly Conference Abstracts 16,* (pp. 4489).

Bibri, S. E. (2015). Implicit and natural HCI in AmI: ambient and multimodal user interfaces, intelligent agents, intelligent behavior, and mental and physical invisibility. In *The Human Face of Ambient Intelligence* (pp. 259-318). Amsterdam: Atlantis Press.

Blesser, B. & Salter, L. R. (2009). *Spaces speak, are you listening?: experiencing aural architecture.* Massachusetts: MIT press.

Bovermann, T., Hermann, T., & Ritter, H. (2006). A tangible environment for ambient data representation. In *First International Workshop on Haptic and Audio Interaction Design, 2.*

Brazil, E. & Fernström, M. (2006). Investigating concurrent auditory icon recognition. In *Proceedings of the 12th International Conference on Auditory Display (ICAD 2006), London, UK, 20-23 June.*

Brazil E. & Fernström M. (2007). Investigating ambient auditory information systems. In *Proceedings of the 13th International Conference on Auditory Display, Montreal, Canada, June26-29, 2007, pp. 326-333,*

Brazil, E. & Fernström, M. (2011). Auditory icons. In Hermann, Tproe., Hunt, A., & Neuhoff, J. G., (Eds), *The Sonification Handbook,* (pp. 325–338). Berlin, Germany: Logos Publishing House.

Bregman, A. S. (1990). *Auditory scene analysis.* Cambridge, MA: MIT Press.

Brier, S. (2008). The paradigm of Peircean biosemiotics. *Signs, 2,* 20-81.

Brooks, R. (2003). *Robot: The Future of Flesh and Machines*. London: Penguin Books.

Brower, C. (2000). A cognitive theory of musical meaning. *Journal of Music Theory, 44*(2), 323-379.

Carlile, S. (2011). Psychoacoustics. In Hermann, T., Hunt, A., & Neuhoff, J. G. (Eds) *The Sonification Handbook*, (pp. 41–61). Berlin, Germany: Logos Publishing House.

Carnap, R. (1928). 1967, *The Logical Structure of the World-Pseudoproblems in Philosophy.* Trans. Rolf A. George. Berkeley: University of California Press. Originally published as Der logische Aufbau der Welt (Berlin-Schlachtensee: Weltkreis).

Cassidy R.J., Berger J., Lee K., Maggioni M., & Coifman R.R. (2004). Auditory display of hyperspectral colon tissue images using vocal synthesis models. In: Barrass S. & Vickers P. (eds) *Proceedings of the 10th International Conference on Auditory Display (ICAD2004), Sydney, Australia, July 6-9, 2004.*

Chafe, C., Caceres, J. P., & Iorga, M. (2013). Sound synthesis for a brain stethoscope. *The Journal of the Acoustical Society of America, 134*(5), 4053-4053.

Chion, M. (1990). *Audio-vision: Sound on Screen,* New York: Columbia University Press. trans. C. Gorbman.

Church, A. (1936). A note on the Entscheidungsproblem. *The Journal of Symbolic Logic, 1*(01), 40-41.

Clarke, J. M. (1992). An FOF synthesis tutorial. In *The CSound Manual; Appendix 4.*

Clarke, M. (2000). FOF and FOG Synthesis in Csound. In *The Csound Book* (pp. 293-306). Massachusetts: MIT Press.

Clark, T. (1987). Echoic memory explored and applied. *Journal of Services Marketing, 1*(2), 41-48.

Cook, P. R. (2002). *Real sound synthesis for interactive applications.* Wellesley: AK Peters.

Cowan N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychological Bulletin, 104,* 163-191.

Cowan N. (1995). *Attention and memory: An integrated framework.* Oxford: Oxford University Press.

Cox, A. (1999). *The metaphoric logic of musical motion and space* (Unpublished Doctoral dissertation). University of Oregon.

Cox, A. (2001). The mimetic hypothesis and embodied musical meaning. *Musicae Scientiae, 5*(2), 195-212.

Cox, C. (2011). Beyond representation and signification: Toward a sonic materialism. *Journal of Visual Culture, 10*(2), 145-161.

Damasio, A. (2008). *Descartes' error: Emotion, reason and the human brain.* Random House.

Darwin, C. J., Turvey, M. T., & Crowder, R. G. (1972). An auditory analogue of the Sperling partial report procedure: Evidence for brief auditory storage. *Cognitive Psychology, 3*(2), 255-267.

Degara, N., Nagel, F., & Hermann, T. (2013). SonEX: an evaluation exchange framework for reproducible sonification. In *Proceedings of the 19th International Conference on Auditory Displays, Lodz, Poland, July 6-10, 2013.*

Dervinytė, I. (2009). Conceptual emigration and immigration metaphors in the language of the press: A contrastive analysis. *Kalbų Studijos,* 4, 49-55.

Descartes, R. (1647). Meditations on First Philosophy in which are demonstrated the existence of God and the distinction between the human soul and body. *Descartes: Selected Philosophical Writings*.

Dewey, J. (1934). *Having an Experience. Art as experience.* New York: Minton, Balch and Co. (pp 36-59).

DeWitt, A., & Bresin, R. (2007, September). Sound design for affective interaction. In *International Conference on Affective Computing and Intelligent Interaction* (pp. 523-533). Springer Berlin Heidelberg.

Diniz N., Deweppe A., Demey M., & Leman M. (2010). A Framework for music-based interactive sonification. In *Proceedings of 16th International Conference on Auditory Display (ICAD 2010), pp. 345-351, Washington DC, USA, June 9-15, 2010.*

Diniz, N., Coussement, P., Deweppe, A., Demey, M., & Leman, M. (2012). An embodied music cognition approach to multilevel interactive sonification. *Journal on Multimodal User Interfaces, 5*(3-4), 211-219.

Dombois, F. & Eckel, G. (2011). Audification. In Hermann, T., Hunt, A., Neuhoff, J. G., (Ed.) *The Sonification Handbook,* (pp. 301–324). Berlin, Germany: Logos Publishing House.

Dourish, P. (2004). *Where the action is: the foundations of embodied interaction.* Massachusetts; MIT press.

Dreyfus, H. L. (1965). *Alchemy and artificial intelligence.* Santa Monica, CA: The Rand Corporation, Research Report.

Dreyfus, H. L. (1992). *What computers still can't do: a critique of artificial reason.* Massachusetts: MIT press.

Droumeva, M., & Wakkary, R. (2008). Understanding aural fluency in auditory display design for ambient intelligent environments. In *Proceedings of the 14th International Conference on Auditory Displays, Paris, France, June 24-27, 2008.*

Droumeva, M., Antle, A.N., Corness, G. & Bevans, A. (2009). Springboard: Exploring embodied metaphor in the design of sound feedback for physical responsive environments. In *Proceedings of the 15th International Conference on Auditory Display, Cophenhagen, Denmark, May 18-22, 2009.*

Droumeva, M., de Castell, S., & Wakkary, R. (2007). Investigating sound intensity gradients as feedback for embodied learning. In *Proceedings of the 13th International Conference for Auditory Display (pp. 26-9), Montreal, Canada, Jun3 26-29, 2007.*

Durand, J. (2005). *Tense/lax, the vowel system of English and phonological theory. Headhood, elements, specification and contrastivity,* (pp 77-98).

Effenberg, A. O. (2004). Using sonification to enhance perception and reproduction accuracy of human movement patterns. In *Proceedings of the International Workshop on Interactive Sonification, Bielefeld 24* (88), 136.

Eliasziw, M., & Donner, A. (1991). Application of the McNemar test to non independent matched pair data. *Statistics in Medicine, 10*(12), 1981-1991.

Elordieta, G. (Ed.) & Prieto, P. (Ed.) (2012). Prosody and meaning. *Encyclopedia of Computer Science.* Berlin, Boston: De Gruyter Mouton. Retrieved 25 Apr. 2015, from http://www.degruyter.com/view/product/177731

Emmerson, S. (2013). *Living electronic music.* Aldershot and Burlington, VT. Ashgate Publishing, Ltd.

Evans, V. & Green, M. (2006). *Cognitive Linguistics: an introduction.* Edinburgh, Scotland: Edinburgh University Press.

Fauconnier, G. & Turner, M. (2002). *The way we think: conceptual blending and the mind's hidden complexities.* New York: Basic Books.

Fergusson, S., Martens, W. L., & Cabrera, D. (2011). Statistical sonification for exploratory data analysis. In Hermann, T., Hunt, A., & Neuhoff, J. G., editors, *The Sonification Handbook*, (pp. 175–196). Berlin, Germany: Logos Publishing House.

Fernström, M., & Brazil, E. (2001). Sonic Browsing: an auditory tool for multimedia asset management. Hiipakka, J, Zacharov, N., & Takala, T., (Eds). In *Proceedings of the 7th International Conference on Auditory Display (ICAD2001)*, *Espoo, Finland, July 29-August 1, 2001.*

Fernström, M., & Brazil, E. (2009). The Shannon Portal: Designing an auditory display for casual users in a public environment. In *Proceedings of the 15th International Conference on Auditory Display*, *Copenhagen, Denmark, 2009,* (pp. 27–30).

Flowers, J. H. (2005). Thirteen years of reflection on auditory graphing: Promises, pitfalls, and potential new directions. *Proceedings of the 11th International Conference on Auditory Display (ICAD2005), Limerick, Ireland, July 6-9, 2005.*

Fodor, J. A. (1975). *The language of thought* (Vol. 5). Harvard University Press.

Fogassi, L., & Ferrari, P. F. (2007). Mirror neurons and the evolution of embodied language. *Current Directions in Psychological Science, 16*(3), 136-141.

Fox, J., & Carlile, J. (2005). SoniMime: movement sonification for real-time timbre shaping. In *Proceedings of the 2005 Conference on New Interfaces for Musical Expression* (pp. 242-243). National University of Singapore.

Franinović, K., & Serafin, S. (2013). *Sonic interaction design.* Massachusetts; MIT Press.

Gallese, V. & Lakoff, G. (2005). The brain's concepts: The role of the sensory-motor system in conceptual knowledge. *Cognitive Neuropsychology, 22*(3-4), 455-479.

Gardner, H. (1985). *The mind's new science.* Basic Books.

Gaver, W. W. (1989). The SonicFinder: An interface that uses auditory icons. *Human–Computer Interaction, 4*(1), 67-94.

Gaver, W. W. (1993). Synthesizing auditory icons. In *Proceedings of the INTERACT'93 and CHI'93 Conference on Human Factors in Computing Systems* (pp. 228-235). ACM.

Gentilucci, M. & Corballis, M. C. (2006). From manual gesture to speech: a gradual transition. *Neuroscience & Biobehavioral Reviews, 30*(7), 949-960.

Gibson, J. J. (1978). The ecological approach to the visual perception of pictures. *Leonardo, 11*(3), 227-235.

Godøy, R. I. (1997). *Knowledge in music theory by shapes of musical objects and sound-producing actions* (pp. 89-102). Springer Berlin: Heidelberg.

Godøy, R. I. (2006). Gestural-sonorous objects: embodied extensions of Schaeffer's conceptual apparatus. *Organised Sound, 11*(02), 149-157.

Grossman, J. (2010). From metaphor to medium: sonification as extension of our body, in *Proceedings of the16th International Conference on Auditory Display (ICAD2010), Washington, D.C, June 9-15, 2010.*

Graham, R. & Bridges, B. (2014). Gesture and embodied metaphor in spatial music performance systems design. In *Proceedings of the International Conference on New Interfaces for Musical Expression (2014)* (pp. 581-584). Goldsmiths University of London.

Graham, R. & Bridges, B. (2015). Managing musical complexity with embodied metaphors. In *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*. Louisiana State University.

Grandjean, D., Sander, D., Pourtois, G., Schwartz, S., Seghier, M. L., Scherer, K. R., & Vuilleumier, P. (2005). The voices of wrath: brain responses to angry prosody in meaningless speech. *Nature Neuroscience, 8*(2), 145-146.

Grieser, D. L. & Kuhl, P. K. (1988). Maternal speech to infants in a tonal language: Support for universal prosodic features in motherese. *Developmental Psychology, 24*(1), 14.

Groehn, M. & Takala, T. (1995). Magicmikes: method for spatial sonification. In *IS&T/SPIE's Symposium on Electronic Imaging: Science & Technology* (pp. 294 301). International Society for Optics and Photonics.

Grond, F. & Berger, J. (2011). Parameter mapping sonification. In Hermann, T., Hunt, A., Neuhoff, J. G., (Eds) *The Sonification Handbook,* (pp. 363–397). Berlin, Germany: Logos Publishing House.

Grond, F., Bovermann, T. & Hermann, T. (2011). A supercollider class for vowel synthesis and its use for sonification. In *The 17th Annual Conference on Auditory Display, Budapest, Hungary 20-24 June, 2011.*

Grond, F. & Hermann, T. (2012). Aesthetic strategies in sonification. *AI & Society*, *27*(2), 213-222.

Grond, F. & Hermann, T. (2014). Interactive Sonification for Data Exploration: How listening modes and display purposes define design guidelines. *Organised Sound, 19* (01), 41-51.

Halpern, A. R., Zatorre, R. J., Bouffard, M. & Johnson, J. A. (2004). Behavioral and neural correlates of perceived and imagined musical timbre. *Neuropsychologia, 42*(9), 1281-1292.

Hampe, B. & Grady, J. E. (Eds.). (2005). *From perception to meaning: Image schemas in cognitive linguistics, 29*. Walter de Gruyter.

Harnad, S. (1990). The symbol grounding problem. *Physica D: Nonlinear Phenomena*, *42*(1), 335-346.

Heidegger, M. (1927). *The Basic Problems of Phenomenology.* Trans. Albert Hofstadter. Bloomington: Indiana University Press.

Heidelberger, M. (2004). *Nature from within: Gustav Theodor Fechner and his psychophysical worldview.* University of Pittsburgh Press.

Hermann, T. (2002). *Sonification for exploratory data analysis.*(Doctoral Dissertation) Retrieved from: http://sonification.de/publications/media/Hermann2002-SFE.pdf

Hermann, T. (2008). Taxonomy and definitions for sonification and auditory display. In *Proceedings of the 14th International Conference on Auditory Display (ICAD2008), Paris, France. 24-27 June, 2008.*

Hermann, T. (2011). Model-based sonification, In Hermann T., Hunt A., and Neuhoff J. G., (Eds) *The Sonification Handbook,* (pp. 399–427). Logos Publishing House [Online]. Available: http://sonification.de/handbook/ chapters/ chapter16/

Hermann T, Baier G, Stephani U. & Ritter H. (2006). Vocal sonification of pathologic EEG features. In: Stockman T (ed) *Proceedings of the 12th International Conference on Auditory Display, London, UK, June 20- 23,* (pp 158–163).

Hermann, T. & Baier, G. (2009). Interpreting EEG by voice: vocal EEG sonification. In M. Rinnot (Ed.), *Proceedings of the 1st Sketching Sonic Interaction Design Workshop Holon, Israel: COST IC0601 SID and HIT.*

Hermann, T., Hildebrandt, T., Langeslag, P. & Rinderle-Ma, S. (2015). Optimizing aesthetics and precision in sonification for peripheral process-monitoring. In *Proceedings of the 21st International Conference on Auditory Display (ICAD 2015). Graz, Austria, July 8-10.*

Hermann, T. & Hunt, A. (2004). The importance of interaction in sonification. In *Proceedings of 10th International Conference on Auditory Display (ICAD 2004),* Sydney, Australia, July 6-9, 2004.

Hermann, T., Hunt, A. & Neuhoff, J. G. (2011). Introduction. In Hermann, T., Hunt, A., Neuhoff, J. G., editors, *The Sonification Handbook,* (pp.1–6) Berlin, Germany; Logos Publishing House.

Hermann, T., Krause, J. & Ritter, H. (2002). Real-time control of sonification models with a haptic interface. In *Proceedings of the 8th International Conference on Auditory Display (ICAD2002), Kyoto, Japan, July 2-5, 2002* (pp. 82-86).

Hermann, T. & Ritter, H. (1999). Listen to your data: Model-based sonification for data analysis. *Advances in Intelligent Computing and Multimedia systems (8),* 189-194.

Hermann, T. & Ritter, H. (2004). Sound and meaning in auditory data display. *Proceedings of the IEEE, 92*(4), 730-741.

Hirschberg, J. (2002). Communication and prosody: Functional aspects of prosody. *Speech Communication, 36*(1), 31-43.

Hodgson, G. M. (1993). The economy as an organism—not a machine. *Futures, 25*(4), 392-403.

Hubbard, T. L. (2010). Auditory imagery: empirical findings. *Psychological Bulletin, 136*(2), 302.

Hunt, A. & Hermann, T. (2011). Interactive sonification. In Hermann, T., Hunt, A., Neuhoff, J. G., editors, *The Sonification Handbook,* (pp 273–298). Logos Publishing House, Berlin, Germany.

Hurtienne, J. & Israel, J.H. (2007). Image schemas and their metaphorical extensions – Intuitive patterns for tangible interaction. In *Proceedings of the 1st International Conference on Tangible and Embedded Interaction 2007* (pp. 127-134), ACM Press.

Husserl, E. (1913).*Ideas: general introduction to pure phenomenology.* Routledge. Originally published 1913. Reprinted 2012.

Imaz, M., & Benyon, D. (2007). *Designing with blends: conceptual foundations of human-computer interaction and software engineering methods.* Massachusetts: MIT Press.

Intons-Peterson, M. J. (1992). Components of auditory imagery. In D. Reisberg (Ed.), *Auditory Imagery* (pp. 45–71). Hillsdale, NJ: Erlbaum.

Johnson, M. (1987). *The body in the mind: the bodily basis of meaning, imagination, and reason.* Chicago: University of Chicago Press.

Johnson, M., (2008) *The meaning of the body: aesthetics of human understanding.* Chicago, IL: University of Chicago Press.

Johnson, M. (2010). Embodied knowing through art. *The Routledge Companion to Research in the Arts,* (pp 141-151).

Johnson, M. (2013). II. Identity, Bodily Meaning, and Art. *Consciousness, Literature & the Arts, 32,* 15.

Johnson, M. L., & Larson, S. (2003). “Something in the way she moves” -Metaphors of musical motion. *Metaphor and Symbol,* 18(2), 63-84.

Johnson, M., & Rohrer, T. (2007). We are live creatures: Embodiment, American Pragmatism, and the cognitive organism. In T. Ziemke, J. Zlatev, & Frank R.(Eds.), *Body, language and mind, Vol. 1: Embodiment.* (pp. 17–54). Amsterdam: Mouton de Gruyter.

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code?. *Psychological Bulletin, 129*(5), 770.

Kabisch, E., Kuester, F., & Penny, S. (2005). Sonic panoramas: experiments with interactive landscape image sonification. In *Proceedings of the 2005 International Conference on Augmented Tele-existence* (pp. 156-163). ACM.

Kahneman, D. (2011). *Thinking, fast and slow.* Dublin: Macmillan.

Kendall, G. S. (2010). Meaning in electroacoustic music and the everyday mind. *Organised Sound, 15*(01), 63-74.

Kendall, G. S. (2014). The Feeling Blend: Feeling and emotion in electroacoustic art. *Organised Sound, 19*(02), 192-202.

Kiefer, M., Sim, E. J., Herrnberger, B., Grothe, J., & Hoenig, K. (2008). The sound of concepts: four markers for a link between auditory and conceptual brain systems. *The Journal of Neuroscience, 28*(47), 12224-12230.

Kilander, F., & Lonnqvist, P. (2002). A whisper in the woods—an ambient soundscape for peripheral awareness of remote processes. In *Proceedings of the 8th International Conference on Auditory Display (ICAD 2002), Kyoto, Japan, July 2-5, 2002.*

Kleinbaum, D. G., & Klein, M. (2010). *Analysis of matched data using logistic Regression* (pp. 389-428). Springer New York.

Klemmer, S. R., Hartmann, B., & Takayama, L. (2006). How bodies matter: five themes for interaction design. In *Proceedings of the 6th Conference on Designing Interactive systems* (pp. 140-149). ACM.

Koestler, A. (1967). *The act of creation: A study of the conscious and unconscious in science and art.* Dell Publishing Company.

Kramer, G. (1993). *Auditory display: Sonification, audification, and auditory interfaces.* Perseus Publishing.

Kramer, G., Walker, B. N., Bonebright, T., Cook, P., Flowers, J., Miner, N. & Tipei, S. (1999). *The sonification report: Status of the field and research agenda.* Report prepared for the national science foundation by members of the international community for auditory display. International Community for Auditory Display (ICAD), Santa Fe, NM.

Kuhn, T. S. (1962). *The structure of scientific revolutions.* Chicago: University of Chicago Press.

Lakoff, G. (1994). *Master metaphor list.* California: University of California.

Lakoff, G. (1987). *Fire, and dangerous things. What categories reveal about the mind.* Chicago: University of Chicago Press.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by.* Chicago: University of Chicago Press.

Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh: The embodied mind and its challenge to western thought.* New York: Basic books.

Lakoff, G., & Núñez, R. E. (2000). *Where mathematics comes from: How the embodied mind brings mathematics into being.* New York: Basic books.

Lancaster, H. O. (1969). *Chi-square distribution.* John Wiley & Sons, Inc..

Larson, S. (2010). *Musical forces: motion, metaphor, and meaning in music.* Bloomington: Indiana University Press.

Leman, M. (2008). *Embodied music cognition and mediation technology.* Massachusetts: MIT Press.

Lighthill, J. (1973). *Report to the Science Research Council on funding for artificial intelligence research.*

Loeb, L. (1981) *From Descartes to Hume: continental metaphysics and the development of modern philosophy.* Ithaca, NY: Cornell University Press.

Macaranas, A., Antle, A. N., & Riecke, B. E. (2012). Bridging the gap: attribute and spatial metaphors for tangible interface design. In *Proceedings of the 6th International Conference on Tangible, Embedded and Embodied Interaction* (pp. 161-168).

Martinez, M., Besold, T. R., Abdel-Fattah, A., Gust, H., Schmidt, M., Krumnack, U., & Kühnberger, K. U. (2012). Theory blending as a framework for creativity in systems for general intelligence. *Theoretical Foundations of Artificial General Intelligence, 4,* 219-239.

Maturana, H. R., & Varela, F. J. (1987). *The tree of knowledge: The biological roots of human understanding.* New Science Library/Shambhala Publications.

Mauney, B. S., & Walker, B. N. (2004). Creating functional and livable soundscapes for peripheral monitoring of dynamic data. In *Proceedings of the 10th International Conference on Auditory Display (ICAD-04),* Sydney, Australia, 2004.

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics, 5*(4), 115-133.

McGookin, D. & Brewster, S. (2011). Earcons. In Hermann, T., Hunt, A., & Neuhoff, J. G. Editors, *The Sonification Handbook*, (pp 339–361). Logos Publishing House, Berlin, Germany.

McNemar, Q. (1955). *Psychological statistics.* (2nd ed.) New York: Wiley.

Meyer, L. B. (1956). *Emotion and meaning in music.* Chicago: University of Chicago Press.

Miller, G. A. (1975). Stanley Smith Stevens. *Biographical Memoirs,* 47, 425.

Nasir, T., & Roberts, J. C. (2007). Sonification of spatial data. In *The 13th International Conference on Auditory Display (ICAD 2007)* (pp. 112-119). ICAD.

Naumann, A., Hurtienne, J., Israel, J. H., Mohs, C., Kindsmüller, M. C., Meyer, H. A., & Hußlein, S. (2007). *Intuitive use of user interfaces: defining a vague concept. In Engineering psychology and cognitive ergonomics* (pp. 128-136). Springer: Berlin Heidelberg.

Nees, M. A. (2009). *Internal representations of auditory frequency: behavioral studies of format and malleability by instructions.* (Unpublished Ph.D. Dissertation). Georgia: Georgia Institute of Technology.

Nees, M. A., & Best, K. (2013). Modality and encoding strategy effects on a verification task with accelerated speech, visual text, and tones. In *Proceedings of the 19th International Conference on Auditory Display (ICAD2013), Lodz, Poland, July 6-9 2013.*

Nees, M. A., & Walker, B. N. (2008). Encoding and representation of information in auditory graphs: Descriptive reports of listener strategies for understanding data. In *Proceedings of the 14th International Conference on Auditory Display, Paris, France, June 24-27, 2008.*

Nees, M. A., & Walker, B. N. (2009). Auditory interfaces and sonification. *The Universal Access Handbook*, 507-521.

Nees, M. A., & Walker, B. N. (2011). Auditory displays for in-vehicle technologies. *Reviews of Human Factors and Ergonomics, 7*(1), 58-99.

Neuhoff, J. G. (2011). Perception, Cognition and Action in Auditory Displays. In Hermann, T., Hunt, A., & Neuhoff, J. G. Editors, *The Sonification Handbook*, pages 63-85. Berlin, Germany: Logos Publishing House.

Neuhoff, J. G., & Heller, L. M. (2005). One small step: Sound sources and events as the basis for auditory graphs. 1st Symposium on Auditory Graphs, in *Proceedings of the 11th International Conference on Auditory Display,* University of Limerick, Limerick, Ireland, July 2005.

Neuhoff, J. G., Knight, R., & Wayand, J. (2002). Pitch change, sonification, and musical expertise: Which way is up? In *Proceedings of the 8th International Conference on Auditory Display*, Kyoto, Japan July 2-5 2002.

Noë, A. (2009). *Out of our heads: Why you are not your brain, and other lessons from the biology of consciousness.* New York, NJ: Hill & Wang.

Núñez, R. & Freeman, W. (1999). *Reclaiming cognition: the primacy of action, intention and emotion*. Imprint Academic.

Oh, J., & Wang, G. (2012). Evaluating crowdsourcing through amazon mechanical turk as a technique for conducting music perception experiments. In *Proceedings of the 12th International Conference on Music Perception and Cognition* (pp. 1-6).

Parseihian, G., & Katz, B. F. (2012). Morphocons: A new sonification concept based on morphological earcons. *Journal of the Audio Engineering Society, 60*(6), 409-418.

Peirce, C. S. (1897/1955). Logic as semiotic: The theory of signs. In Buchler, J. (ed.), *The Philosophical Writings of Peirce,* Dover, New York, (pp. 98–119).

Peres, S.C., & Lane, D.M. (2005). Auditory graphs: The effects of redundant dimensions and divided attention. *Proceedings of the 11th International Conference on Auditory Display (ICAD2005)* (pp.169-174), Limerick, Ireland.

Peres, S. C. (2012). A comparison of sound dimensions for auditory graphs: Pitch is not so perfect. *Journal of the Audio Engineering Society, 60*(7/8), 561-567.

Polli, A. (2012). Soundscape, sonification, and sound activism. *AI & Society, 27*(2), 257-268.

Polotti, P., Delle Monache, S., Papetti, S., & Rocchesso, D. (2008, April). Gamelunch: forging a dining experience through sound. In *CHI'08 Extended Abstracts on Human Factors in Computing Systems* (pp. 2281-2286). ACM.

Ponty, M. M. (1968). *The visible and the invisible. Maurice Merleau-Ponty: Basic Writings.* Psychology Press.

Popper, K. R. (1970). *Normal science and its dangers* (pp. 57-58). Cambridge: Cambridge University Press.

Putnam, H. (1967). Psychological predicates. *Art, mind, and religion, (1)*, 37-48.

Quinn, M.(2001). Research set to music: The climate symphony and other sonifications of ice core, radar, DNA, seismic, and solar wind data. In *Proceedings of the 7th International Conference on Auditory Display (ICAD2001),* Espoo, Finland.

Ribes-Iñesta, E. (2003). What is defined in operational definitions? The case of operant psychology. *Behavior and Philosophy,* 111-126.

Rocchesso, D., & Fontana, F. (2003). *The Sounding Object.* Mondo Estremo.

Rocchesso, D., & Bresin, R. (2007). Emerging sounds for disappearing computers. In *The disappearing computer* (pp. 233-254). Springer: Berlin Heidelberg.

Rocchesso, D., Polotti, P., & Delle Monache, S. (2009). Designing continuous sonic interaction. *International Journal of Design, 3*(3), 13-25.

Roddy, S., & Furlong, D. (2013a). Embodied cognition in auditory display. In *Proceedings Of the 19th International Conference on Auditory Display,* July 6-10 Łódź, Poland (pp. 127-134).

Roddy, S., & Furlong, D. (2013b). Rethinking the Transmission Medium in Live Computer Music Performance. Presented at the *Irish Sound Science and Technology Convocation*, Dún Laoghaire Institute of Art, Design and Technology.

Roddy, S., & Furlong, D. (2014). Embodied Aesthetics in Auditory Display. *Organised Sound, 19*(01), 70-77.

Roddy, S., & Furlong, D. (2015a). Embodied Auditory Display Affordances. In *Proceedings of the 12th Sound and Music Computing Conference:* National University of Ireland Maynooth (pp. 477-484).

Roddy, S., & Furlong, D. (2015b). Sonification Listening: An Empirical Embodied Approach. In *Proceedings of The 21st International Conference on Auditory Display (ICAD 2015)* July 8–10, 2015, Graz, Austria. (pp.181-187).

Rodet, X., Potard, Y., & Barriere, J. B. (1984). The CHANT project: From the synthesis of the singing voice to synthesis in general. *Computer Music Journal*, *8*(3), 15-31.

Rimland, J., Ballora, M., & Shumaker, W. (2013). Beyond visualization of big data: a multi-stage data exploration approach using visualization, sonification, and storification. In *SPIE Defense, Security, and Sensing* (pp. 87580K-87580K). International Society for Optics and Photonics.

Rumsey, F. (2001). *Spatial Audio Focal.* Oxford, U.K.: Focal.

Ryle, G. (1949). *The concept of mind.* London, UK: Hutchinson.

Santa Ana, O. (1999). ‘Like an animal I was treated’: Anti-immigrant metaphor in US public discourse. *Discourse & Society*, *10*(2), 191-224.

Sándor, A., & Lane, D. M. (2003). Sonification of absolute values with single and multiple dimensions. In *Proceedings of the 9th International Conference on Auditory Display (ICAD2003),* Boston, MA, July 7-9, 2003.

Scaletti, C. (1994). Auditory display: sonification, audification, and auditory interfaces. *Sound Synthesis Algorithms for Auditory Data Representations, 1,* 223-251.

Schaeffer, P. (1966). *In search of a concrete music.* Trans. John Dack, Christine North. Los Angeles: University of California Press.

Schafer, R. M. (1969). *The new soundscape: a handbook for the modern music teacher.* Canada: BMI.

Scherer, K. R. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin, 99,* 143–165.

Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences, 3*(03), 417-424.

Serafin, S., Franinovic, K., Hermann, T., Lemaitre, G., Rinott, M., & Rocchesso, D. (2011). Sonic interaction ′ design. In Hermann, T., Hunt, A., Neuhoff, J. G., editors, *The Sonification Handbook,*(pp. 87–110). Berlin, Germany: Logos Publishing House.

Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of information.* Urbana, Illinois: University of Illinois Press.

Shepard, R. N. (1964). Circularity in judgments of relative pitch. The *Journal of the Acoustical Society of America, 36*(12), 2346-2353.

Smalley, D. (1986). Spectro-morphology and structuring processes. *The Language of Electroacoustic Music,* 61-93.

Smalley, D. (1996). The listening imagination: listening in the electroacoustic era. *Contemporary Music Review, 13*(2), 77-107.

Smalley, D. (1997). Spectromorphology: explaining sound-shapes. *Organised Sound, 2*(02), 107-126.

Smith, D. R., & Walker, B. N. (2005). Effects of auditory context cues and training on performance of a point estimation sonification task. *Applied Cognitive Psychology, 19*(8), 1065-1087.

Spondike, D.(2006). A Strategy for Composing Music using the Sonification of "Snapshot" Type Data Collections “Schnappschuss von der Erde” . In *Proceedings of the 12th International Conference on Auditory Display,* London, UK, June 20-23, 2006.

Stevens, S. S. (1975). *Psychophysics.* Transaction Publishers.

Taylor, S., & Fernström, M. (2012). Marbh Chrios. *Leonardo, 45*(2), 192-193.

Todes, S. (2001). *Body and World.* Cambridge, MA: The MIT Press.

Truax, B. (1978). *Handbook for Acoustic Ecology,* Originally published by the World Soundscape Project, Simon Fraser University.

Truax, B. (1984). *Acoustic communication.* Norwood, NJ: Ablex Publishing Corporation.

Truax, B. (1996). Soundscape, acoustic communication and environmental sound composition. *Contemporary Music Review, 15*(1-2), 49-65.

Truax, B. (2012). Sound, listening and place: The aesthetic dilemma. *Organised Sound, 17*(03), 193-201.

Turing, A. M. (1936). On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society,* Series 2, 41:230–267.

Turing, A. M. (1947). The state of the art. vortrag vor der londoner mathematical society am 20. februar 1947. Alan M. Turing, *Intelligence Service.* Schriften hrsg. von, 2.

Turing, A. M. (1950). Computing machinery and intelligence. *Mind,* 433-460.

Vaissière, J. (1983). Language-independent prosodic features. In *Prosody: Models and measurements* (pp. 53-66). Berlin Heidelberg: Springer.

van Zuijen, T. L., Sussman, E., Winkler, I., Näätänen, R., & Tervaniemi, M. (2005). Auditory organization of sound sequences by a temporal or numerical regularity—a mismatch negativity study comparing musicians and non-musicians. *Cognitive Brain Research, 23*(2), 270-276.

Varela, F.J., Thompson, E.T. & Rosch, E. (1991). *The Embodied Mind: Cognitive Science and Human Experience.* Cambridge, MA: The MIT Press.

Vercoe, B. Ekman, D. B., Conder, K., Yi, S., Gogins, M., & Cabrera, A.(2015). *The Canonical Csound Reference Manual.* (605 ed.). Cambridge MA: MIT Press.

Vermersch, P. (2009). Describing the practice of introspection. *Journal of Consciousness Studies, 16*(10-12), 20-57.

Vickers, P. (2005). Ars Informatica -- Ars Electronica: Improving Sonification Aesthetics. In: *Understanding and Designing for Aesthetic Experience* (workshop at HCI 2005: The 19th British HCI Group Annual Conference), 5-9 September 2005, Edinburgh.

Vickers, P. (2011). Sonification for process monitoring. In Hermann, T., Hunt, A., & Neuhoff, J. G., (Eds). *The Sonification Handbook,* (pp. 455–491). Berlin, Germany: Logos Publishing House.

Vickers, P. (2012). Ways of listening and modes of being: electroacoustic auditory display. *Journal of Sonic Studies, 2*(1), 1-23.

Vickers, P., & Hogg, B. (2006).Sonification abstraite/sonification concrete: An ‘aesthetic perspective space’ for classifying auditory displays in the ars musica domain, in *Proceedings of the 12th International Conference on Auditory Display*, London, UK, June 2006.

Wakkary, R., Hatala, M., Lovell, R., & Droumeva, M. (2005). An ambient intelligence platform for physical play. In *Proceedings of the 13th annual ACM international conference on Multimedia* (pp. 764-773).

Walker, B. N. (2002). Magnitude estimation of conceptual data dimensions for use in sonification. *Journal of Experimental Psychology: Applied, 8*(4), 211.

Walker, B. N., & Kramer, G. (2004). Ecological psychoacoustics and auditory displays: Hearing, grouping, and meaning making. *Ecological Psychoacoustics,* 150-175.

Walker, B. N., & Kramer, G. (2005). Mappings and metaphors in auditory displays: An experimental assessment. *ACM Transactions on Applied Perception (TAP), 2*(4), 407-412.

Walker, B.N. & D.M. Lane. (2001). “Psychophysical scaling of sonification mappings: A comparison of visually impaired and sighted listeners,” in *Proceedings 7th of the International Conference on Auditory Display.* Espoo, Finland.

Walker, B. N. & Nees, M. A. (2011). Theory of sonification. In Hermann, T., Hunt, A., Neuhoff, J. G., editors, *The Sonification Handbook,* pages 9–39. Berlin, Germany: Logos Publishing House.

Watson, R. (2007). *Cogito, ergo sum: The life of René Descartes.* David R. Godine Publisher.

Watson, R. A. (1993). Shadow history in philosophy. *Journal of the History of Philosophy, 31*(1), 95-109.

Wiener, N. (1948). *Cybernetics, or Communication and Control in the Animal and the Machine.* New York: Wiley.

Westerkamp, H. (1999). Soundscape composition: Linking inner and outer worlds. *Soundscape Newsletter,* Amsterdam, Holland.

Westerkamp, H. (2007). The world soundscape project. *World,* 12, 41.

White, M. (2003). Metaphor and economics: the case of growth. *English for Specific Purposes, 22*(2), 131-151.

Whitehead, A. N., & Russell, B. (1912). *Principia mathematica* (Vol. 2). Cambridge: Cambridge University Press.

Wiener, N. (1961). *Cybernetics or control and communication in the animal and the machine (Vol. 25).* Massachusetts: MIT press.

Wishart, T. (1996). *On sonic art. A new and revised edition.* Amsterdam: Harwood Academic Publishers.

Wittgenstein, L. (1921). *Tractatus logico-philosophicus. Annalen der Naturphilosophie.* With second English translation by DF Pears and BF McGuiness.

Wolf, K. E., Gliner, G., & Fiebrink, R. (2015, March). A model for data-driven sonification using soundscapes. In *Proceedings of the 20th International Conference on Intelligent User Interfaces Companion* (pp. 97-100). ACM.

Worrall, D. (2009). *Sonification and information: Concepts, instruments and techniques.* Ph.D. thesis. University of Canberra.

Worrall, D. (2010). Parameter mapping sonic articulation and the perceiving body. In *Proceeding of the 16th International Conference on Auditory Display,* Washington, D.C, USA, June 9-15, 2010.

Worrall, D. (2011). A Method for Developing an Improved Mapping Model for Data Sonification. In *Proceedings of the 17th International Conference on Auditory Display, Budapest, June 20-24, 2011,* (pp. 20-24).

Worrall, D. (2013). Understanding the need for microgestural inflections in parameter mapping sonification. In *Proceedings of the 19th International Conference on Auditory Display.* Lodz, Poland.

Zbikowski, L. M. (1997). Conceptual models and cross-domain mapping: New perspectives on theories of music and hierarchy. *Journal of Music Theory,* 193-225.

Zbikowski, L.M., (2005). *Conceptualizing music: Cognitive Structure, Theory, and Analysis.* New York, US: Oxford University Press, 2005.

Żołnowska, I. (2011). Weather as the source domain for metaphorical expressions. *AVANT.*Pismo Awangardy Filozoficzno-Naukowej, (1), 165-179.

APPENDICES

Appendix 2.1: Example Evaluation Questions for Chapter 2

Evaluation B. Q 7 to 12

41) The following clip represents the Number of ants in an ant colony over a 1-year period.
Does the number of ants Increase or Decrease over time?*

Increase Decrease

42) The following clip represents changes in average Depth along the course of a river.
Does the river become Deeper or Shallower along its course?*

Deeper Shallower

43) The following clip represents the Number of students attending a school over a 3-year period.
Does the number of students Increase or Decrease over time?*

Increase Decrease

44) The following clip represents the size of the average annual Corn Yield harvested in Ireland over a 10-year period.
Does the average Yield get Larger or Smaller over time?*

Larger Smaller

45) The following clip represents the Temperature of a specific location over a 3-week period.
Does the Temperature get Hotter or Colder over time?*

Hotter Colder

46) The following clip represents the Size of benign tumor over a 3-year period.
Does the tumor get Bigger or Smaller over time?*

Bigger Smaller

Appendix 2.2: Demographic Data for Chapter 2 Evaluation

Gender	Percent		Count
Male	74.10%		83
Female	25.90%		29
Age	**Percent**		**Count**
under 18	1.80%		2
18-24	26.80%		30
25-34	33.90%		38
35-54	33.90%		38
55+	3.60%		4
Nationality	**Percent**		**Count**
Bosnia and Herzegovina	6.30%		7
Bulgaria	6.30%		7
India	7.10%		8
Indonesia	3.60%		4
Italy	5.40%		6
Netherlands	3.60%		4
Poland	3.60%		4
Portugal	3.60%		4
Serbia	8.00%		9
Spain	8.00%		9
Turkey	6.30%		7
Venezuela	4.50%		5
Algeria	1.80%		2
Argentina	0.90%		1
Brazil	1.80%		2
Canada	2.70%		3
Colombia	1.80%		2
Egypt	0.90%		1
Estonia	0.90%		1
France	0.90%		1
Greece	1.80%		2
Hungary	2.70%		3
Macedonia	1.80%		2
Mexico	0.90%		1
Morocco	1.80%		2
Pakistan	0.90%		1
Peru	1.80%		2

Philippines	0.90%		1
Romania	1.80%		2
Russia	1.80%		2
Slovakia	1.80%		2
Sweden	0.90%		1
Ukraine	1.80%		2
United States	1.80%		2
Formal Music Training	**Percent**		**Count**
Yes	25.90%		29
No	74.10%		83
Years of Formal Training	**Percent**		**Count**
Less than 1 year	44.80%		13
3 years or less	37.90%		11
5 years or more	17.20%		5
Play an Instrument	**Percent**		**Count**
Yes	22.30%		25
No	77.70%		87
Instrumental Skill Level	**Percent**		**Count**
Basic	60.00%		15
Intermediate	36.00%		9
Advanced	4.00%		1

Appendix 3.1: Example Evaluation Questions for Chapter 3

Speed Evaluation

Please listen to Audio 1 and rate it on the following scale from FAST to SLOW*
Very Slow Slow Medium Fast Very Fast

Please listen to Audio 2 and rate it on the following scale from FAST to SLOW*
Very Slow Slow Medium Fast Very Fast

Please listen to Audio 3 and rate it on the following scale from FAST to SLOW*
Very Slow Slow Medium Fast Very Fast

Please listen to Audio 4 and rate it on the following scale from FAST to SLOW*
Very Slow Slow Medium Fast Very Fast

Please listen to Audio 5 and rate it on the following scale from FAST to SLOW*
Very Slow Slow Medium Fast Very Fast

Please listen to Audio 6 and rate it on the following scale from FAST to SLOW*
Very Slow Slow Medium Fast Very Fast

Please listen to Audio 7 and rate it on the following scale from FAST to SLOW*
Very Slow Slow Medium Fast Very Fast

Please listen to Audio 8 and rate it on the following scale from FAST to SLOW*
Very Slow Slow Medium Fast Very Fast

Please listen to Audio 9 and rate it on the following scale from FAST to SLOW*
Very Slow Slow Medium Fast Very Fast

Please listen to Audio 10 and rate it on the following scale from FAST to SLOW*
Very Slow Slow Medium Fast Very Fast

Please listen to Audio 11 and rate it on the following scale from FAST to SLOW*
Very Slow Slow Medium Fast Very Fast

Please listen to Audio 12 and rate it on the following scale from FAST to SLOW*
Very Slow Slow Medium Fast Very Fast

Timbral Evaluation 1

Please listen to Audio 25 and rate the sound using the pairs of descriptions below.

Rate this sound in terms of Size*
Big Small

Rate this sound in terms of Brightness*
Dark Bright

Rate this sound in terms of Weight*
Heavy Light

Rate this sound in terms of Strength*
Strong Weak

Rate this sound*
Rough Smooth

Rate this sound*
Hot Cold

Emotional Parameter Evaluation
Please listen to Audio 32 and choose an emotion that the sound invokes?*
Tenderness Happiness Anger Sadness Fear

Please listen to Audio 33 and choose an emotion that the sound invokes?*
Tenderness Happiness Anger Sadness Fear

Please listen to Audio 34 and choose an emotion that the sound invokes?*
Tenderness Happiness Anger Sadness Fear

Please listen to Audio 35 and choose an emotion that the sound invokes?*
Tenderness Happiness Anger Sadness Fear

Please listen to Audio 36 and choose an emotion that the sound invokes?*
Tenderness Happiness Anger Sadness Fear

Appendix 3.2: Demographic Data for Chapter 3 Evaluation.

Gender	**Percent**		**Count**
Male	74.10%		103
Female	25.90%		36
Age	**Percent**		**Count**
under 18	2.20%		3
18-24	15.80%		22
25-34	43.20%		60
35-54	33.80%		47
55+	5.00%		7
Nationality	**Percent**		**Count**
Bosnia and Herzegovina	4.30%		6
Bulgaria	7.90%		11
India	6.50%		9
Italy	5.80%		8
Poland	3.60%		5
Romania	5.80%		8
Russia	3.60%		5
Serbia	3.60%		5
Spain	10.80%		15
Venezuela	7.20%		10
Afghanistan	0.70%		1
Albania	0.70%		1
Algeria	2.20%		3
Argentina	0.70%		1
Brazil	0.70%		1
Canada	2.20%		3
Egypt	0.70%		1
Finland	1.40%		2
Germany	1.40%		2
Greece	2.90%		4
Hong Kong	0.70%		1
Hungary	0.70%		1
Indonesia	1.40%		2
Ireland	0.70%		1
Jordan	1.40%		2
Macedonia	0.70%		1
Malaysia	0.70%		1

Mexico	0.70%		1
Netherlands	2.20%		3
Pakistan	2.20%		3
Peru	1.40%		2
Philippines	1.40%		2
Portugal	2.20%		3
Qatar	0.70%		1
Sweden	1.40%		2
Tunisia	1.40%		2
Turkey	2.20%		3
Ukraine	0.70%		1
United Kingdom	1.40%		2
United States	0.70%		1
Vietnam	2.20%		3
Formal Music Training	**Percent**		**Count**
Yes	20.10%		28
No	79.90%		111
Years of Formal Training	**Percent**		**Count**
Less than 1 year	71.90%		100
3 years or less	15.80%		22
5 years or more	7.20%		10
More than 8 years	5.00%		7
Play an Instrument	**Percent**		**Count**
Yes	26.60%		37
No	73.40%		102
Instrumental Skill Level	**Percent**		**Count**
None	31.70%		44
Basic	37.40%		52
Intermediate	24.50%		34
Advanced	6.50%		9

Appendix 3.3: 2nd Prototyping Platform

2nd Prototyping Platform

```
<CsoundSynthesizer>
<CsOptions>
; Select audio/midi flags here according to platform
-odac   ;;;RT audio out
;-iadc  ;;;uncomment -iadc if RT audio input is needed too
; For Non-realtime ouput leave only the line below:
; -o comb.wav -W ;;; for file output any platform
</CsOptions>
<CsInstruments>

sr = 44100
ksmps = 32
nchnls = 2
0dbfs = 1

/****USER DEFINED OPCODES****/

opcode vowgen2,       a, kkkkkki

kfund,kmorf,koct,imode,kris,kdur,kdec xin  ;Opcode extended by Stephen Roddy to
accept grain params

imorf   ftgentmp 0, 0, 16, 10, 1; must be 16 elements long because vowels are in tables of
length 16
ifenv   ftgentmp 0, 0, 4096, 19, .5, .5, 270, .5
ivib    ftgentmp 0, 0, 4096, 10, 1
if imode == 0 igoto bass
if imode == 1 igoto tenor
if imode == 2 igoto countertenor
if imode == 3 igoto alto
if imode == 4 igoto soprano
bass:
ia      ftgentmp 0, 0, 16, -2, 600, 1040, 2250, 2450, 2750, 0, -7, -9, -9, -20, 60, 70, 110,
120, 130
ie      ftgentmp 0, 0, 16, -2, 400, 1620, 2400, 2800, 3100, 0, -12, -9, -12, -18, 40, 80, 100,
120, 120
ii      ftgentmp 0, 0, 16, -2, 350, 1700, 2700, 3700, 4950, 0, -20, -30, -22, -28, 60, 90,
100, 120, 120
io      ftgentmp 0, 0, 16, -2, 450, 800, 2830, 3500, 4950, 0, -11, -21, -20, -40, 40, 80, 100,
120, 120
iu      ftgentmp 0, 0, 16, -2, 325, 700, 2530, 3500, 4950, 0, -20, -32, -28, -36, 40, 80, 100,
120, 120
igoto ind
tenor:
```

```
ia      ftgentmp 0, 0, 16, -2, 650, 1080, 2650, 2900, 3250, 0, -6, -7, -8, -22, 80, 90, 120,
130, 140
ie      ftgentmp 0, 0, 16, -2, 400, 1700, 2600, 3200, 3580, 0, -14, -12, -14, -20, 70, 80,
100, 120, 120
ii      ftgentmp 0, 0, 16, -2, 290, 1870, 2800, 3250, 3540, 0, -15, -18, -20, -30, 40, 90,
100, 120, 120
io      ftgentmp 0, 0, 16, -2, 400, 800, 2600, 2800, 3000, 0, -10, -12, -12, -26, 70, 80, 100,
130, 135
iu      ftgentmp 0, 0, 16, -2, 350, 600, 2700, 2900, 3300, 0, -20, -17, -14, -26, 40, 60, 100,
120, 120
igoto ind
countertenor:
ia      ftgentmp 990, 0, 16, -2, 660, 1120, 2750, 3000, 3350, 0, -6, -23, -24, -38, 80, 90,
120, 130, 140
ie      ftgentmp 991, 0, 16, -2, 440, 1800, 2700, 3000, 3300, 0, -14, -18, -20, -20, 70, 80,
100, 120, 120
ii      ftgentmp 992, 0, 16, -2, 270, 1850, 2900, 3350, 3590, 0, -24, -24, -36, -36, 40, 90,
100, 120, 120
io      ftgentmp 993, 0, 16, -2, 430, 820, 2700, 3000, 3300, 0, -10, -26, -22, -34, 40, 80,
100, 120, 120
iu      ftgentmp 994, 0, 16, -2, 370, 630, 2750, 3000, 3400, 0, -20, -23, -30, -34, 40, 60,
100, 120, 120
igoto ind
alto:
ia      ftgentmp 0, 0, 16, -2, 800, 1150, 2800, 3500, 4950, 0, -4, -20, -36, -60, 80, 90, 120,
130, 140
ie      ftgentmp 0, 0, 16, -2, 400, 1600, 2700, 3300, 4950, 0, -24, -30, -35, -60, 60, 80,
120, 150, 200
ii      ftgentmp 0, 0, 16, -2, 350, 1700, 2700, 3700, 4950, 0, -20, -30, -36, -60, 50, 100,
120, 150, 200
io      ftgentmp 0, 0, 16, -2, 450, 800, 2830, 3500, 4950, 0, -9, -16, -28, -55, 70, 80, 100,
130, 135
iu      ftgentmp 0, 0, 16, -2, 325, 700, 2530, 3500, 4950, 0, -12, -30, -40, -64, 50, 60, 170,
180, 200
igoto ind
soprano:
ia      ftgentmp 0, 0, 16, -2, 800, 1150, 2900, 3900, 4950, 0, -6, -32, -20, -50, 80, 90, 120,
130, 140
ie      ftgentmp 0, 0, 16, -2, 350, 2000, 2800, 3600, 4950, 0, -20, -15, -40, -56, 60, 100,
120, 150, 200
ii      ftgentmp 0, 0, 16, -2, 270, 2140, 2950, 3900, 4950, 0, -12, -26, -26, -44, 60, 90,
100, 120, 120
io      ftgentmp 0, 0, 16, -2, 450, 800, 2830, 3800, 4950, 0, -11, -22, -22, -50, 40, 80, 100,
120, 120
iu      ftgentmp 0, 0, 16, -2, 325, 700, 2700, 3800, 4950, 0, -16, -35, -40, -60, 50, 60, 170,
180, 200
igoto ind

ind:
index   ftgentmp 0, 0, 16, -2, ia, ie, ii, ia, io, iu, ie, io, ii, iu, ia, io, ia, ia, ia, ia, ia
```

```
	ftmorf	kmorf, index, imorf

kfx	=	0
kform1	table	kfx,  imorf
kform2	table	kfx+1, imorf
kform3	table	kfx+2, imorf
kform4	table	kfx+3, imorf
kform5	table	kfx+4, imorf
kamp1	table	kfx+5, imorf
kamp2	table	kfx+6, imorf
kamp3	table	kfx+7, imorf
kamp4	table	kfx+8, imorf
kamp5	table	kfx+9, imorf
kbw1	table	kfx+10,imorf
kbw2	table	kfx+11,imorf
kbw3	table	kfx+12, imorf
kbw4	table	kfx+13, imorf
kbw5	table	kfx+14, imorf

iolaps	=	100
a1	fof	ampdbfs(kamp1), kfund, kform1, koct, kbw1, kris,  kdur,  kdec, iolaps,
ivib, ifenv, p3
a2	fof	ampdbfs(kamp2), kfund, kform2, koct, kbw2, kris,  kdur,  kdec, iolaps,
ivib, ifenv, p3
a3	fof	ampdbfs(kamp3), kfund, kform3, koct, kbw3, kris,  kdur,  kdec, iolaps,
ivib, ifenv, p3
a4	fof	ampdbfs(kamp4), kfund, kform4, koct, kbw4, kris,  kdur,  kdec, iolaps,
ivib, ifenv, p3
a5	fof	ampdbfs(kamp5), kfund, kform5, koct, kbw5, kris,  kdur,  kdec, iolaps,
ivib, ifenv, p3
asig	=	a1+a2+a3+a4+a5

	xout	asig
endop

/****USER DEFINED OPCODES****/


/***********************DATA SET STORAGE***********************/


;DATA SET INPUT
							;Gross National Product from 2007-
2012 by quarter
	giGNP ftgen 0, 0, -24, -2,
37283,36546,37108,37091,37448,36956,35934,34948,34127,33147,32506,32227,32367,3
2894,33574,33887,32820,32749,32716,32360,32597,33958,33094,33339,33339
```

```
;DATA                                        ;Avg. Unemployment Rate from 2007-2012 by quarter
          giUER ftgen 0, 0, -24, -2,       4.5, 4.66, 4.63,4.86, 5.06, 5.7, 6.86, 8.13, 10.2, 11.96, 12.66, 13.03, 13.1, 13.66, 14, 14.63, 14.43,14.43, 14.73, 14.9, 15, 14.76, 14.63, 14.23, 14.23


;DATA                                        ;Avg. Emigration Rate from 2007-2012 in thousands

  giEMI ftgen 0, 0, -6, -2,          46.3, 49.2, 72.0, 69.2, 80.6, 87.1      , 87.1


;DATA
                                                            ;Data
                    giTEMP ftgen 0, 0, -14, -2, 8.5,8.0,8.9,12.6,15.1,19.3,22.1,19.4,18.7,20,22,25,31

instr 1

gkfund =148 ;GLOBAL FUNDAMENTAL VARIABLE


 /***************DATA PROCESSING****************/

iDatHi = 37448
iDatLo = 32227

          iDatReadRate = p3
          iDatLen ftlen giGNP
          kDatRead line 0, iDatReadRate, iDatLen
          kDatIn tablei kDatRead,  giGNP
          ;**DATA SCALING**
          ky = (((1)*(kDatIn - iDatLo))/(iDatHi - iDatLo) ); 0-1

 /****************PROSODY********************/

          ;Basic Carrier Lfo defines phrase/prosody shape
kSlfo =10

kSpeed = 2
k2      randh 1, 3   ;Randomisation to emulation phrasing and prosody
kScps = kSpeed*k2
aPhrase  lfo kSlfo, kScps, 0



          ;Up Lfo models + rate
kUlfo =5  ;Control the Amount of LFO applied to the signal here
```

```
kUSpeed = .5
kU2     randh 1, .3   ;Randomisation to emulation phrasing and prosody
kUcps = kUSpeed*kU2
aUp  lfo kUlfo, kUcps, 4



          ;Down Lfo models - rate
kDlfo =5  ;Control the Amount of LFO applied to the signal here

kDSpeed = .5
kD2     randh 1, .3   ;Randomisation to emulation phrasing and prosody
kDcps = kDSpeed*kD2
aDn  lfo kDlfo, kDcps, 5

/****************PausingModel*********************/

          ;Square Wave LFO models pauses
kPlfo =1  ;Control the Amount of LFO applied to the signal here
kPSpeed random .25,2
   ;Randomisation to emulation phrasing and prosody

kPR2   randh 1, .25

kP2 randomh .0625,10,.5


kPcps = kPSpeed*kPR2


aPs  lfo kPlfo, kPcps, 3


          ;Add Up and Down Signal to Phrase Signal Signals Together
aProsody = (aPhrase + aUp + aDn)

koct =0
;printk 0,ky
/****************AMBIENT VOICE PROCESSIG****************/
kmorf1= 7-(4*ky) ; the polarity is switched because GNP is a falling value

;0 1 2 3 4 5 6 7 8 9 10 11 12
; a e I a o u e o I u a o a

/****************PROCESS VOCAL TIMBRAL
PARAMATERS****************/

kris =.003
kdur = .002
```

```
kdec = .007


ktone = 1.25*ky
kRndLfoTim randomi 1,6,3

kFLfo lfo 1,kRndLfoTim
kFLfo = kFLfo + 2
kBaseFund = gkfund*ktone
kminfund = 60

kfund = kBaseFund ;+ kFLfo

kfund = (kminfund +kfund+aProsody); * aPs


/***********************VOWGEN AREA****************************/
 ;asig vowgen kfreq,  kmorf, koct, kris, kdur, kdec, imode
 aVox1 vowgen2 kfund, kmorf1, koct, kris, kdur, kdec,  3
 aVox2 vowgen2 kfund, kmorf1, koct, kris, kdur, kdec,  4
 aVox3 vowgen2 kfund, kmorf1, koct, kris, kdur, kdec,  5

 aVox = (aVox1+aVox2+aVox3)/3 ;Keep the Avox Signal Useable

 /******************************************************************/

  aVox butterlp aVox, 3000

 asigl= aVox
 asigr= aVox

;aoutL, aoutR freeverb ainL, ainR, kRoomSize, kHFDamp[, iSRate[, iSkip]]


aL,   aR freeverb asigl, asigr,  .01,   .1, sr

  /* Post-Processing**/


  asigl= asigl*.15
  asigr= asigr*.15

  aoutL = asigl+aL
  aoutR= asigr+aR

  aL1, aR1 pan2 aoutL, 1-ky,0
  aL2, aR2 pan2 aoutR, 1-ky,0

  aleft = aL1+aL2
  aright = aR1+aR2
```

```
  outs aleft, aright
   endin

/************************************************************/

instr 2 ;UNEMPLOYMENT
 gkfund =148

/* IN ORDER TO SONIFY THE DATA MUST FIRST ANALYSE IT FOR HIGH AND
LOW POINTS |AND ENTER HERE*/
/*************************************************************/

          ;DATA PROCESSING
iDatHi = 15
iDatLo = 4.5

          iDatReadRate = p3
          iDatLen ftlen giUER
          kDatRead line 0, iDatReadRate, iDatLen
          kDatIn tablei kDatRead, giUER


          ;**DATA SCALING**
          ky = (((1)*(kDatIn - iDatLo))/(iDatHi - iDatLo)); 0-1




          printk 0,ky
/*************************************************************/

/****************** CHOIR MODELING ******************/
;0 1 2 3 4 5 6 7 8 9 10 11 12
;a e I a o u e o I u a o a

koct =0

/******************AMBIENT VOICE PROCESSIG******************/
kmorf1= 4+(2.25*ky) ; starts at 3(a) and moves signal up another 3 to reach 6(e)

;0 1 2 3 4 5 6 7 8 9 10 11 12
; a e I a o u e o I u a o a

/****************** PROCESS VOCAL TIMRAL
PARAMATERS******************/
kris =.003
kdur = .002
kdec = .007
```

```
kRndLfoTim randomi 1,6,3

ktone = .25*ky

kFLfo lfo 1,kRndLfoTim
kFLfo = kFLfo + 2
kBaseFund = (gkfund)*ktone ;Keeps the Fundamental frequencies relative to one another!
kfund = kBaseFund + kFLfo

/***************** PROCESS VOCAL TIMRAL
PARAMATERS*****************/


/************************VOWGEN AREA*****************************/
 ;asig vowgen kfreq,  kmorf, koct, kris, kdur, kdec, imode

 aVox1 vowgen2 60+kfund, kmorf1, koct, kris, kdur, kdec,  0
 aVox2 vowgen2 60+kfund, kmorf1, koct, kris, kdur, kdec, 1
 aVox3 vowgen2 60+kfund, kmorf1, koct, kris, kdur, kdec, 2


 aVox =(aVox1+aVox2+aVox3)/3 ;Keep the Avox Signal Useable

/************************************************************************/

  aVox butterlp aVox, 3000

 asigl= aVox
 asigr= aVox

/********************************REVERB*********************************
*/
;aoutL, aoutR freeverb ainL, ainR, kRoomSize, kHFDamp[, iSRate[, iSkip]]
aL,   aR freeverb asigl, asigr,  .01,   .1, sr

  /* Post-Processing**/


  asigl= asigl*.1
  asigr= asigr*.1

  aoutL = asigl+aL
  aoutR= asigr+aR

  aL1, aR1 pan2 aoutL, 1-ky,0
  aL2, aR2 pan2 aoutR, 1-ky,0

  aleft = aL1+aL2
  aright = aR1+aR2
```

```
 outs aleft, aright
/**********************************************************************/

  endin

/**********************************************************************/

instr 3 ;EMIGRATION

 ;GLOBAL VARIABLES
 gkfund =148

/* IN ORDER TO SONIFY THE DATA MUST FIRST ANALYSE IT FOR HIGH AND
LOW POINTS |AND ENTER HERE*/
/**********************************************************************/

        ;DATA PROCESSING
iDatHi = 87.1
iDatLo = 46.3

        iDatReadRate = p3
        iDatLen ftlen giEMI
        kDatRead line 0, iDatReadRate, iDatLen
        kDatIn tablei kDatRead, giEMI
        ;**DATA SCALING**



        ky = ((1)*(kDatIn - iDatLo))/(iDatHi - iDatLo) ; 0-1

printk 0,ky
 /**********************************************************************/


/***********************CHOIR MODELING***********************/
;0 1 2 3 4 5 6 7 8 9 10 11 12
;a e I a o u e o I u a o a

koct =0

/***********************AMBIENT VOICE
PROCESSIG***********************/
kmorf1= (1*ky) ; starts at 3(a) and moves signal up another 3 to reach 6(e)

;0 1 2 3 4 5 6 7 8 9 10 11 12
; a e I a o u e o I u a o a

/***********************PROCESS VOCAL TIMRAL
PARAMATERS***********************/
kris =.003
```

```
kdur = .002
kdec = .007

kRndLfoTim randomi 1,6,3

ktone = .125*ky

kFLfo lfo 1,kRndLfoTim
kFLfo = kFLfo + 2
kBAseFund = (gkfund)*ktone ;Keeps the Fundamental frequencies relative to one
another!
kfund = kBAseFund + kFLfo
/*********************** PROCESS VOCAL TIMRAL
PARAMATERS**********************/


/**********************VOWGEN AREA**********************/
 ;asig vowgen kfreq,  kmorf, koct, kris, kdur, kdec, imode

 aVox1 vowgen2 60+kfund, kmorf1, koct, kris, kdur, kdec,  0
 aVox2 vowgen2 60+kfund, kmorf1, koct, kris, kdur, kdec, 1
 aVox3 vowgen2 60+kfund, kmorf1, koct, kris, kdur, kdec, 2


 aVox =(aVox1+aVox2+aVox3)/3 ;Keep the Avox Signal Useable

/********************************************************************/

  aVox butterlp aVox, 3000

 asigl= aVox
 asigr= aVox

;aoutL, aoutR freeverb ainL, ainR, kRoomSize, kHFDamp[, iSRate[, iSkip]]
aL,   aR freeverb asigl, asigr,  .01,   .1, sr

/*********************** Post-Processing **********************/


  asigl= asigl*.1
  asigr= asigr*.1

  aoutL = asigl+aL
  aoutR= asigr+aR

  aL1, aR1 pan2 aoutL, .5+(.5*ky),0
  aL2, aR2 pan2 aoutR, .5-+(.5*ky),0

  aleft = aL1+aL2
  aright = aR1+aR2
```

```
  outs aleft, aright

   endin

/***********************************************/


</CsInstruments>
<CsScore>
; sine wave
f 1 0 4096 10 1
; sigmoid wave
f 2 0 1024 19 0.5 0.5 270 0.5


i1 0 30
i2 0 30
i3 0 30

e

</CsScore>
</CsoundSynthesizer>
```

Appendix 3.4: Compositional Process for The Human Cost and Development of Hypotheses

The technical aspects of the compositional process for *The Human Cost* are discussed in section 3.10 of this thesis. The creative and artistic aspects of the process, which in tandem with exploration with the prototyping platforms, resulted in the development of the hypotheses listed in section 3.11, are considered in this appendix. This appendix describes the findings upon which these hypotheses were based and from which the experimental stimuli, provided in the digital appendices and described in this appendix, were derived.

The starting point for the compositional process for *The Human Cost* began with a consideration of the voice that was motivated by Smalley's statement that "In electroacoustic music the voice always announces a human presence" (Smalley, 1996). If this is true then the human voice might prove effective across a number of levels for the representation of data sets that measure the lives and experiences of real people in an auditory display context. On this basis the choice was made to focus on socioeconomic data sets from the period of Ireland's economic crash and recession. Deprivation, unemployment, emigration and GNP rates in Ireland from 2007 to 2012 were chosen as data sets for the sonification.

Exploration with the prototyping platforms was performed alongside the compositional process. These explorations consisted of attempting to model embodied schemata discussed across the embodied cognition literature using the prototyping platforms. They bore fruit in a number of interesting ways. The earliest explorations were undertaken with the first prototyping platform and focused on controlling the perceived number of vocal gestures and their spatial locations in an auditory display in an attempt to understand how spatial schemata and multitude schemata might be modelled. Both of these proved simple to implement, as spatial processing and the addition of multiple audio sources in the context of a stereo spatial field are well-explored areas of audio research. The chorusing technique discussed in Chapter 3 was also considered as a way of simulating an increase in the perceived number of vocal gestures. A comparison of the addition of multiple vocal gestures (dubbing) to the chorusing technique is presented in the form of an empirical listener evaluation in Chapter 3.

Spatial processing was implemented in *The Human Cost* using panning algorithms, which moved each of the vocal gestures around the stereo space with a rate and directionality determined by the data. The exact locations of the vocal gestures was not important here rather the speed of their motion which was of interest. As the increases and decreases in the data become more pronounced the speed at which they move follows suit amplifying the sense of frenzy in the piece. The spatial dimensionality of the vocal gesture seemed a useful mechanism in the context of sonification. As such a number of simple spatial transformations of vocal gestures that can be accomplished using stereo audio were empirically evaluated in Chapter 3.

The application of a system of low frequency oscillators (lfos) to the output of the fof opcodes in the first prototyping platform suggested that it may be possible to control the perceived speed of a vocal gesture by organising it into phrasal passages and manipulating the rate of change of pitch contours within those passages. *The Human Cost* used a similar process. The output of a random number generator modulated by the emigration data was in turn used to augment the fundamental pitches of a set of vowgen2 fof generators. This helped to humanise the vocal gestures representing emigration and tie it to a sense of wailing or lamentation, methods of singing historically associated with the emigration of the Irish diaspora. It was decided that breaking a vocal gesture into units and manipulating the amplitude envelopes and the speed of presentation of each unit might be a better method for applications beyond the context of the piece.

Many of the outputs of the vowgen2 opcodes sounded as though they were originating from an unrealistically small vocal source. To rectify this an exploration of the size schema might be applied to a vocal gesture were undertaken. These activities suggested that it might be possible to control the perceived size of a vocal gesture by controlling the relationship between amplitude levels, amplitude envelopes, spectral content, pitch and vowel formant profile. Creative post-processing proved fruitful also and the further application of reverb modelling, dynamics modelling via compression, limiting, gating, subtractive stereo imaging and algorithmic loudness maximisation, allowed for the control of perceived in a more exaggerated manner. These parameters were evaluated empirically. For *The Human Cost* the addition of reverbs with simulated room size and high frequency dampening mapped to the data source for each vocal gesture allowed for a subtle modulation of size on the basis of data changes.

During the exploratory process it was also discovered that the vowel formant profiles might also prove useful for controlling the perceived sense of strength in a vocal gesture. Strong-Weak is an attribute schema mentioned repeatedly across the literature. As such, it was decided that the relationship between vowel formant profile and a number of similar attribute schemas should be explored through empirical evaluation. It quickly became obvious that the rough-smooth schema could not be modelled using vowel shape alone. As such different strategies for adding noise to the signal created by the fof generators of the second prototyping platform were explored and two such strategies were selected for empirical testing in Chapter 3. None of these attribute-schema based approaches were employed directly in the human cost. Explorations with vowel formant profiles did open another interesting avenue, as they seemed useful for controlling the sense of tension in a synthesized vocal gesture.

The deprivation rate, unemployment rate and emigration rate in *The Human Cost* were mapped to vowel formant profile so that these economic indicators increased the perceived tension in the vocal gestures would increase also, morphing from an open ‘a’ sound to the more pointed ‘e’. This was found to be subtle yet effective and it was decided to focus on the representation of tension using vowel formant profiles through empirical testing in Chapter 3.

Listening to *The Human Cost* seemed to rouse a specific emotional response. This suggested that it might also be possible to control perceived emotion expressed in a vocal gesture. The representation of emotion using synthesised vocalisations is a complex topic of study. As such a meta-analysis of the area of emotional communication in speech and singing produced by Juslin and Laukka (Juslin and Laukka, 2003) was consulted to provide vocal characteristics of emotional vocal gestures could be operationalised as synthesis parameters within the second prototyping platform. These parameters are listed in the appendices for chapter 3.

Differentiating the three vocal gesture lines used in the human cost proved a challenge at first. In order to make them distinct their timbres had to be manipulated by defining unique grain envelopes for each vowgen2 opcode. This coupled with their parameter-mappings from different data sets helped to lend them distinct identities.

To avoid over cluttering the piece and to introduce a rhythmic element the GNP data was mapped to modulate a sound that was analogous to a heartbeat. The sound of a heartbeat was chosen because as discussed in chapter 3, economies are often conceptualised in terms of living organisms. GNP was mapped to modulate the speed of that heartbeat, and as GNP falls the heartbeat falls signalling the death of the organism, or economy.

The outputs of each of the individual Csound instrument were loaded into Logic Pro X project file for further processing and mixing. Here a system of reverbs and delays was applied to each of the outputs. This system created a dream like texture intended to represent the prosperity of the Celtic Tiger era. As the piece progresses this dreamlike texture is faded out leaving the raw vocal gestures and heartbeat rhythm which signal a nightmarish new reality of frenzied uncertainty.

Appendix 3.5: Detailed Synthesis Parameters

Vocal Gesture: Speed

The sounds for this experiment were developed using additive synthesis techniques. The output of a sine bank of parallel sine oscillators generating 320 partials is passed to a formant filter that further filters the signal to simulate human vowel sounds. These vowel transformations can be modulated to create vocal like transitions between vowel simulations. The audio signal is then multiplied by a control signal to shape an amplitude envelope in terms of the classic attack, decay, sustain and release model. A small amount of reverb is added to enhance the salience of the sound.
For Pitched samples the sine bank outputs a waveform with randomly assigned but statically held partial amplitudes over a central frequency of 440hz.
For Noise samples the sine bank outputs a waveform with randomly assigned and dynamically changing partial amplitudes, creating in essence filtered noise.
The vowel movements are rendered on a scale from “A” at one extreme to “E” at the other by way of “O” in the middle.
The Cross-Modal Acoustic Cues (Juslin and Laukka 2003) relevant to this experiment were hypothesised to be Speech Rate/Tempo and Voice Onset/Tone Attacks as these deal with speed. The synthesis parameters hypothesised to be of relevance to these cues were Note Amounts, Note Lengths, Pause Lengths, Vowel Change Speeds and Amplitude Envelopes.
The following tables show how synthesis parameters were related to these cues.

Slow Parameters

Cross-Modal Acoustic Cues	**Synthesis Parameters**
Speech Rate/Tempo	Fewest Notes Long Note Lengths Largest Pause Between Notes Very Slow (Noise) to No (Pitch) Vowel Changes*
Voice Onset/Tone Attacks	Attack: Long Decay: Medium Sustain: Long Release: Long

Slow Some Parameters

Cross-Modal Acoustic Cues	**Synthesis Parameters**
Speech Rate/Tempo	Few Notes Long Note Lengths Short Pauses Between Notes Quick Vowel Changes*
Voice Onset/Tone Attacks	Attack: Short Decay: Medium Sustain: Long Release: Long

Slow Random Parameters

Cross-Modal Acoustic Cues	**Synthesis Parameters**
Speech Rate/Tempo	Many Notes Mixed Note Lengths Short Pauses Between Notes Quick Vowel Changes*
Voice Onset/Tone Attacks	Attack: Short Decay: Medium Sustain: Long Release: Long

Fast Parameters

Cross-Modal Acoustic Cues	**Synthesis Parameters**
Speech Rate/Tempo	Many Notes Short Note Lengths Short Pauses Between Notes Quick Vowel Changes*
Voice Onset/Tone Attacks	Attack: Short Decay: Medium Sustain: Short Release: Short

Fast Some Parameters

Cross-Modal Acoustic Cues	**Synthesis Parameters**
Speech Rate/Tempo	Fewer Notes Short Note Lengths Long Pauses Between Notes Quick Vowel Changes*
Voice Onset/Tone Attacks	Attack: Long Decay: Medium Sustain: Short Release: Short

Fast Random Parameters

Cross-Modal Acoustic Cues	**Synthesis Parameters**
Speech Rate/Tempo	Fewest Notes Mixed Note Lengths Long and Short Pauses Between Notes Quick Vowel Changes*
Voice Onset/Tone Attacks	Attack: Long Decay: Medium Sustain: Short Release: Short

Vocal Gesture: Size

The sounds for this experiment were also developed using additive synthesis techniques. The output of a sine bank of parallel sine oscillators generating 320 partials. The output of oscillator one is multiplied with a formant filter. For Big level cues that filter has a lowpass profile which boosts frequency at −26 semitones below the root note. For Small level cues that filter has a highpass profile which boosts frequency at +50 semitones above the root note.

For Pitched samples the sine bank outputs a waveform with randomly assigned but statically held partial amplitudes over a central frequency of 27.5hz for pitched BigA0Pitch.wav, 55hz for BigA1Pitch.wav and SmallA1Pitch.wav and 880hz for SmallA5Pitch.wav.

For Noise samples the sine bank outputs a waveform with randomly assigned and dynamically changing partial amplitudes, creating in essence filtered noise. The noise is then filtered (formant filter) to match a pitch profile. As such the noise tones are not pure noise, but rather noisy pitched sounds. BigA0Noise.wav has a central pitch of 27.5hz. BigA1Noise.wav and SmallA1PNoise.wav have a pitch of 55hz and SmallA5Noise.wav has 880hz.

Each signal is passed to a formant filter that simulates human vowel sounds. Sounds with Big level cues are filtered to an “a” vowel profile and sounds with Small level cues are filtered to an “I” vowel sound. Big level signal are then passed to a reverb module that simulates reverb by modifying the amplitudes of the partials rather than using delay lines. For Big level cues the reverb has a Wet Level of −5db, an attack of 58ms, decay of 35ms and hi frequency dampening factor of 38%. Small level cues are not passed to the reverb. Each audio signal is then multiplied by a control signal to shape an amplitude envelope in terms of the classic attack, decay, sustain and release model. The only difference between the parameters for Small A1 and Small A5 are the pitch level. For Big level cues this signal is passed to another reverb unit. The room response for a large arena is applied with 100ms of delay, 0.0db early reflections and 2 second decay times for reverb tails. The reverb response is given maximum stereo spread and cross-mixing (left channel reverb appearing in the right channel and vice versa) is enabled. 100% of the wet signal is added to the original dry signal. The signal is then fed through a 4 band compression module that applies hard compression to even out the energy below 120hz and boost the energy across the rest of the spectrum to provide a more homogenous amplitude profile. The signal is then passed to an imaging module that uses multi-channel delay modelling across 4 bands to widen the image width as much as possible across the entire spectrum. The signal is at last passed through a loudness maximiser that aggressively limits the signal to further maximise volume. Small level signals are passed directly to the multi-channel imaging module in order to narrow the image as much as possible within the centre of the stereo stage.

The following tables show how synthesis parameters were related to these cues.

Big Parameters

Cross-Modal Acoustic Dimensions	**Synthesis Parameters**
Voice Intensity/Sound Level	Lower
High-Frequency Energy	Low Frequency Energy
F0/Pitch Level (fundamental Frequency)	Low
Voice Onset/Tone Attacks	Slow
Vowel Profile	**Vowel Profile**
"a"	"a"
Audio Processing Dimensions	
Pre-Amplitude Envelope Reverb	Large Amount of Reverb
Post-Amplitude Envelope Reverb	Large Amount of Reverb
Dynamics	Large amount of Compression, Limiting and Gating
Stereo Imaging	Large Spread across Stereo Stage
Loudness Maximisation	Large Amount

Small Parameters

Cross-Modal Acoustic Dimensions	**Synthesis Parameters**
Voice Intensity/Sound Level	High
High-Frequency Energy	High Frequency Energy
F0/Pitch Level (fundamental Frequency)	High
Voice Onset/Tone Attacks	Fast
Vowel Profile	**Vowel Profile**
"I"	"I"
Audio Processing Dimensions	
Pre-Amplitude Envelope Reverb	None
Post-Amplitude Envelope Reverb	None
Dynamics	No Compression, Limiting or Gating
Stereo Imaging	Narrow Image in centre of Stereo Stage
Loudness Maximisation	None

Vocal Gestures: Amount

The sounds for this experiment were developed using additive synthesis techniques. The output of a sine bank of parallel sine oscillators generating 320 partials and this homogenous signal is filtered using two formant filters to create a voice like timbre. This signal is filtered with a further formant filter to create to simulate the sound of an "a" vowel. The many pole of the schema is modelled using Logic Pro X's Ensemble plugin. This allows for the application of 8 chorusing effects to a single input signal. These are modulated by two standard LFO's at 100% intensity each with rates of .1 and .2hz respectively.

Vocal Gesture: Attribute Schemata via Vowel Shapes

The sounds for this experiment were developed using additive synthesis techniques. The output of a sine bank of parallel sine oscillators generating 320 partials is passed to a formant filter that further filters the signal to simulate human vowel sounds. The vowels used are a,e,i,o,u,ä and ü . Two versions of each vowel sound, one pitched and one noise, are created totalling fourteen clips in all. The pitched versions have a frequency of 55hz (A1) and the noise versions are created by randomising the amplitudes of the sine 320 partials.

Vocal Gesture: Emotional Dimensions

Evaluation 5 aimed to determine how well listeners could identify emotions in synthesised vocal like tones that were modelled on the basis of Juslin and Laukka's (2003) Cross-Modal Patterns of Acoustic Cues for Discrete Emotions.

All of the Stimuli were modelled using the Razor additive synthesis instrument running in Reaktor 5 which itself is run as a plugin in Logic Pro X. The following sections describe how each of the Cross-Modal Patterns of Acoustic Cues for Discrete Emotions can be were modelled using Razor.

<u>Happiness</u>

Fast speech rate/tempo:
    Quick changes between Vowels in the Vowel Filter
    Quick succession of notes determined in Logics midi piano roll
Medium–high voice intensity/sound level:
    Good volume
    No reverb
Medium high-frequency energy:
    Modulate the signal by a second formant oscillator with a high frequency formant.
    Use a High Pass formant shape equalised to ensure high frequency content.
    Slope bandpass filters to include more high frequency content.
High F0/pitch level:
    Set the MIDI notes to A3
Much F0/pitch variability:
    Pitch modulation by envelope 2 in Razor
Rising F0/pitch contour:
    Drive Envelope 2's pitch modulation upwards
Fast voice onsets/tone attacks:
    Quick attack Razors amp envelope
Very little microstructural irregularity:
    Don't add any noise to the signal.

<u>Sadness</u>

Slow speech rate/tempo:
    Slow Vowel Filter changes
    Single Midi Note

Low voice intensity/sound level:
  Low volume
  Add Reverb to Signal with Ozone 5 Reverb Module
Little voice intensity/sound level variability:
  Keep a standard volume
Little high-frequency energy:
  Use a lowpass Formant shape on Oscillator 2
  Equalise Formant Oscillator to reduce High Frequency Energy
  Shift the Vowel Formant downwards
  Slope Vowel Formant to reduce high frequency energy
  Cap high frequency energy using spectral clip
Low F0/pitch level:
  E1 MIDI note
Little F0/pitch variability:
  Small amount of envelope modulation on pitch
Falling F0/pitch contour:
  Envelope the pitch contour to drop downwards
Slow voice onsets/tone attacks:
  Long envelope attack
Microstructural regularity
  Add some Noise to main oscillator and some amplitude randomisation to oscillator one and two.

Fear

Fast speech rate/tempo:
  Rapid Vowel Filter changes
  Rapid midi notes
Low voice intensity/sound level (except in panic fear):
  Low Volume
  High Reverb amount
Much voice intensity/sound level variability:
  Randomise amplitudes of both oscillators
Little high-frequency energy:
  Use a lowpass Formant shape on Oscillator 2
  Equalise Formant Oscillator to reduce High Frequency Energy
  Shift the Vowel Formant downwards
  Slope Vowel Formant to reduce high frequency energy
  Cap high frequency energy using spectral clip
High F0/pitch level:
  A3 MIDI note
Little F0/pitch variability:
  Stability around the pitch centre.
Rising F0/pitch contour:
  Envelope to drive up pitch contour
Lot of microstructural irregularity
  Add large amount of Noise to main oscillator and some amplitude randomisation to oscillator one and two.

Anger

Fast speech rate/tempo:
- Rapid Vowel Filter changes
- Rapid midi notes

High voice intensity/sound Level
- High volume
- Low amount of reverb

Much voice intensity/sound level variability:
- Randomise amplitudes of both oscillators

Much high-frequency energy:
- Oscillator Two Formant with High Pass shape, Upward Formant Shift, Equalised to emphasise High Frequency Energy
- Vowel Formant Filter with Upward Formant Shift, Slope Vowel Formant to increase high frequency energy
- Set the Singer type to the highest
- Use Ozone EQ to boost the high frequency energy

High F0/pitch level:
- Set the notes to A3

Much F0/pitch variability:
- Plenty of pitch modulation by envelope 2 in Razor

Rising F0/pitch contour:
- Drive Envelope 2's pitch modulation upwards

Fast voice onsets/tone attacks:
- Quick attack on the amp envelope

Microstructural irregularity:
- Very large amount of Noise in Oscillator one.

Tenderness

Slow speech rate/tempo,
- Slow changes across vowels
- One Long midi note

Low voice intensity/sound level
- Low volume level
- Added reverb

Little voice intensity/sound level variability,

- Keep a standard volume

Little high-frequency energy,
- Lowpass Formant shape op Oscillator two, equalised to remove high frequency energy
- Downward shift on the Vowel Formant in Vowel Formant Filter and slope the bandpass filters to factor out higher frequencies.
- Use spectral clip to remove high frequency energy

Low F0/pitch level,
E1 pitch
Little F0/pitch variability,
No lfo on the pitch
Falling F0/pitch contours,
Envelope the pitch contour to drop downwards
Slow voice onsets/tone attacks,
Long envelope attack
Microstructural regularity
No Noise

Vocal Gesture: Spatial Dimensions
All sounds were created in Logic Pro X using the Razor additive synthesis instrument in a Reaktor 5 plugin. The stimuli were all created using pitched noise filtered through a randomly modulated vowel filter which produced vocal like tones. Doppler effects were applied using the Waves Doppler plugin. The pitch range used for the Up-Down schema ranges from is F5 (698hz) to E2. Panning is used to position sounds in the left or right speaker and to move them from one speaker to another. Reverb is used to model distance. For the near pole of the schema a low predelay on the early reflection profile of a smaller acoustic space, high amplitude early reflections, lower overall decay times with higher low frequency decay and lower high frequency decays and a low wet against a high dry level was used. For the far pole of the schema the opposite settings were used. A Long predelay on the early reflection profile of a large arena with low amplitude early reflections, a long decay time, long low and shorter high decay times and a high wet against a low dry level were used.

Vocal Gesture: Tension
This evaluation was intended to determine which vowel sounds are perceived as most and least tense. The evaluation used the exact stimuli as evaluation four, the vowel attribute evaluation.

Vocal Gesture: Texture
The materials were designed using the Razor instrument for Reaktor 5 running as a plugin in the Logic Pro X DAW.
For the Smooth sample the sine bank outputs a waveform with randomly assigned but statically held partial amplitudes over a central frequency of A1 (440hz) determined by a MIDI message.
For the Rough 1 sample the sine bank outputs a waveform with randomly assigned and dynamically changing partial amplitudes, to create a noisy profile, over a central frequency of A1 (440hz) determined by a MIDI message. All samples are filtered to the profile of an “o” vowel.
For the Rough 2 sample the sine bank outputs a waveform with randomly assigned but statically held partial amplitudes over a central frequency of A1 (440hz) determined by a MIDI message. All samples are filtered to the profile of an “o” vowel. Noise is added by applying fast modulation to the partial amplitudes after the vowel formant has been applied to create a beating effect that results in a more noisy sounding timbre.

Appendix 4.1: Example Evaluation Questions for Chapter 4

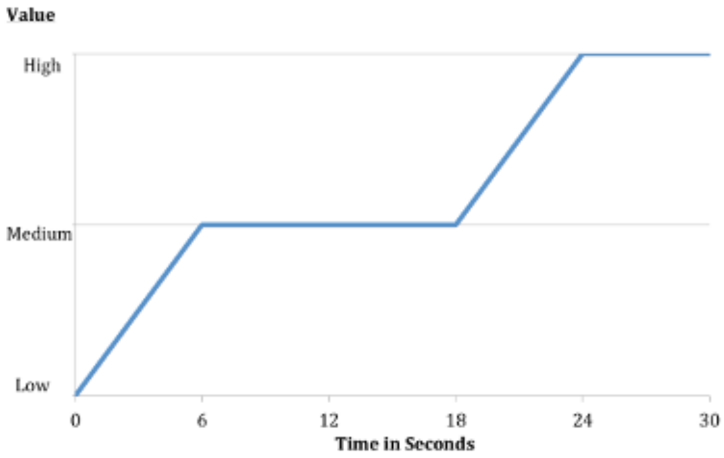
Unemployment

The following clip expresses changes in the Unemployment rate data collected during the Irish Financial Crisis.
Unemployment is expressed through the volume of construction work being carried out on a “Building-Site”.

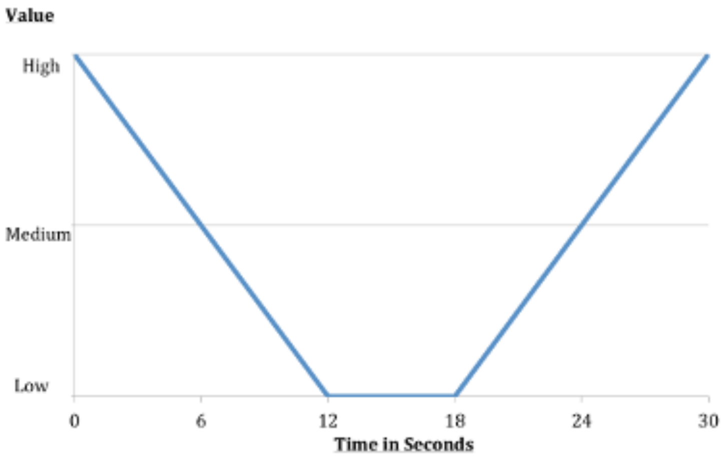
When there is a large degree of Unemployment the building-site sounds abandoned and empty.
When there is a small degree of Unemployment the building-site sounds alive with the sounds of construction work.

Please listen to the clip carefully and answer the questions that follow.
You may replay the clip as many times as you need to help you answer the questions.

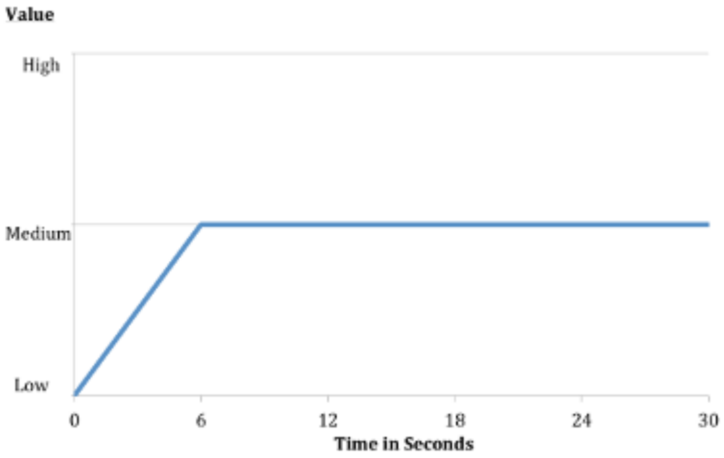
31) How does the above clip express INCREASES and DECREASES in Unemployment rate over time?*



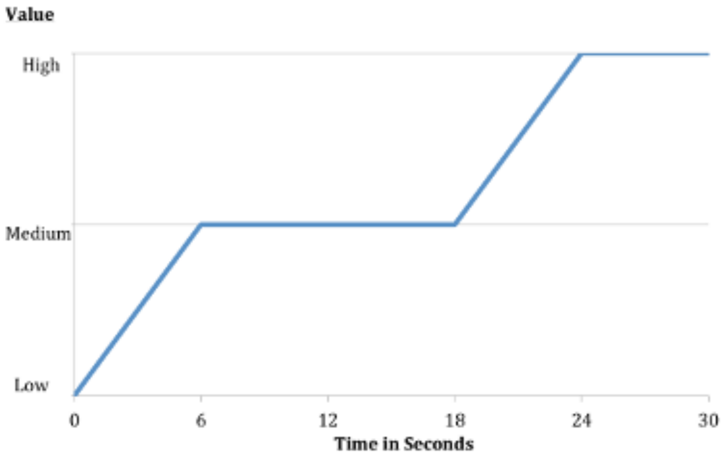
1


Value
High
Medium
Low
0
6
12
18
24
30
Time in Seconds

2


Value
High
Medium
Low
0
6
12
18
24
30
Time in Seconds

3


Value
High
Medium
Low
0
6
12
18
24
30
Time in Seconds

4

32) Are the changes in the Unemployment rate presented as POSITIVE, NEUTRAL or NEGATIVE?*

Positive Neutral Negative

33) What does the clip say about the impact of changes in the Unemployment rate on the SOCIAL situation of those affected?
The impact has been:*

Very Negative Negative No Impact Positive Very Positive

34) What does the clip say about the impact of changes in the Unemployment rate on the MENTAL and EMOTIONAL lives of those affected?
The impact has been:*

Very Negative Negative No Impact Positive Very Positive

35) Rate the effectiveness of the clip in communicating the changes in the Unemployment rate data.*

Very Bad Bad Fair Good Very Good

36) Rate the effectiveness of the clip in communicating the impact of the Unemployment rate changes on those affected.*

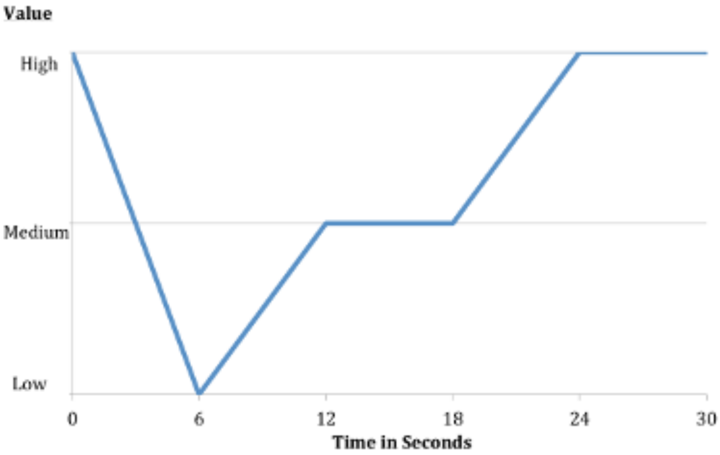
Very Bad Bad Fair Good Very Good

The following clip expresses changes in the UNEMPLOYMENT rate data collected during the Irish Financial Crisis.
Unemployment is expressed through the metaphors of “War” and “Peace”.

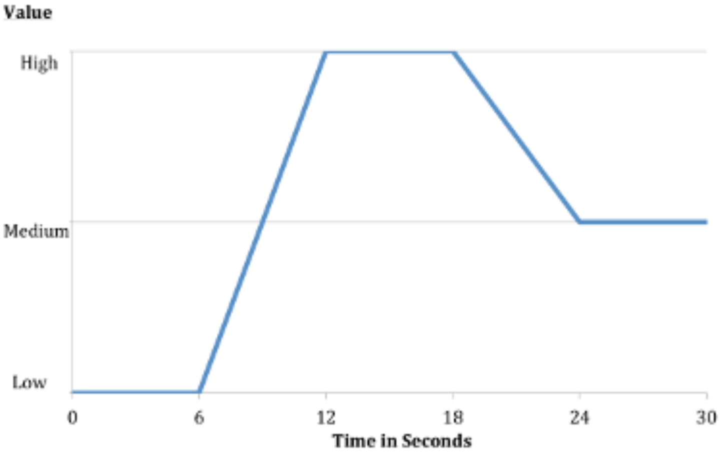
When there is a large degree of Unemployment the clip becomes a War-Zone.
When there is a small degree of Unemployment the clip becomes a Peaceful Zone.

Please listen to the clip carefully and answer the questions that follow.
You may replay the clip as many times as you need to help you answer the questions.

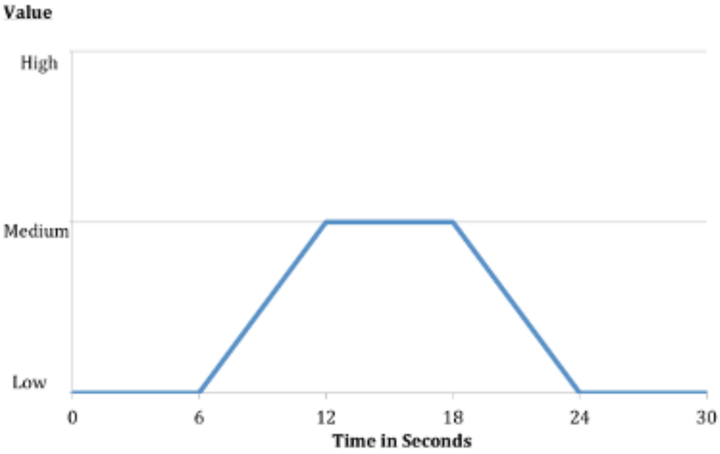
37) How does the above clip express INCREASES and DECREASES in Unemployment over time?*


Value
High
Medium
Low
0
6
12
18
24
30
Time in Seconds

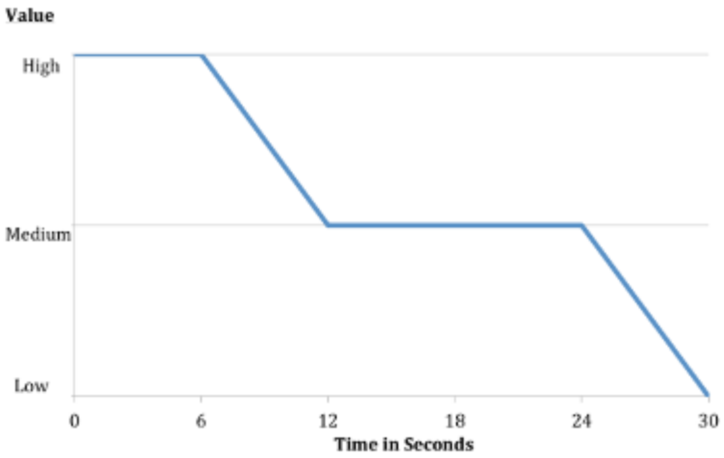
1


Value
High
Medium
Low
0
6
12
18
24
30
Time in Seconds

2


Value
High
Medium
Low
0
6
12
18
24
30
Time in Seconds

3



4

38) Are the changes in the Unemployment rate presented as POSITIVE, NEUTRAL or NEGATIVE?*

Positive Neutral Negative

39) What does the clip say about the impact of changes in the Unemployment rate on the SOCIAL situation of those affected?
The impact has been:*

Very Negative Negative No Impact Positive Very Positive

40) What does the clip say about the impact of changes in the Unemployment rate on the MENTAL and EMOTIONAL lives of those affected?
The impact has been:*

Very Negative Negative No Impact Positive Very Positive

41) Rate the effectiveness of the clip in communicating the changes in the Unemployment rate data.*

Very Bad Bad Fair Good Very Good

42) Rate the effectiveness of the clip in communicating the impact of the Unemployment rate changes on those affected.*

Very Bad Bad Fair Good Very Good

Consider:

An abandoned building site means Unemployment.

A busy building site means Employment.

Listen to the audio below and answer the questions that follow:

43) Does the clip express a Large, Medium or Small amount of Unemployment?*

Large Medium Small

44) Choose the EMOTIONAL aspects of Unemployment highlighted in the clip. You can choose more than one*

Happiness Excitement Tenderness Fear Anger Sadness

45) Does the clip express a Large, Medium or Small amount of Unemployment?*

Large Medium Small

46) Choose the EMOTIONAL aspects of Unemployment highlighted in the clip. You can choose more than one*

Happiness Excitement Tenderness Fear Anger Sadness

47) Does the clip express a Large, Medium or Small amount of Unemployment?*

Large Medium Small

48) Choose the EMOTIONAL aspects of Unemployment highlighted in the clip. You can choose more than one*

Happiness Excitement Tenderness Fear Anger Sadness

Appendix 4.2: Compositional Process for Idle Hands

An overview of the compositional process for *Idle Hands* along with some technical notes is provided in section 4.4. The artistic aspects of the piece are discussed in greater detail in this appendix. The artistic goal of *Idle Hands* composition was to reveal the human dimensions in the Central Statistics Office unemployment rate record from 1983-2014 by linking it to patterns of commonly shared embodied experience through the medium of music. The research goal was to explore and uncover different organisational strategies for data driven soundscape composition on the basis of embodied patterns of experience that might be submitted to empirical listener evaluation later. A compositional approach informed by soundscape studies and soundscape compositional practices were applied.

All fixed sonifications and musical pieces have a beginning middle and end that reflect the source-path-goal structure discussed by Johnson (1987). *Idle Hands* aimed to exploit this inherent embodied schematic structure to represent the time-series. It compressed the passage of time form 1983 to 2014 into 11 minutes and 46 seconds. This approach allows time to represent itself. Flowers (2005) argues that approaches of this nature are effective ways of representing time series data in sound.

Elements of Schaefer's (Schaefer, 1969) conceptual apparatus, the sound signal, soundmark and keynote were employed to organise this sonic content into a soundscape. A sound signal is any sound that a listener is meant to pay attention to in a given soundscape. A soundmark is a sonic analogy to a landmark. It is a unique sound that is significant and specific to a certain community. Keynotes are those omni-present sounds against which have faded into the background of a listener's consciousness and against which foreground signals and soundmarks are perceived and understood.

Building on the compositional techniques employed for *The Human Cost*, simulated vocal sounds might provide a useful source of sonic content for *Idle Hands*. The sonic content chosen for the piece was intended to represent a conceptual blend of the concepts of the human voice, Schaefer's soundscape elements and unemployment. The piece was intended to reflect the effects of unemployment on an imagined human landscape. This choice was motivated by the artistic goal of the piece, which was to reveal the human dimensions of unemployment. Most of the sonic content was created in Csound. The grain opcode was used to create a grain cloud from a sample that was originally generated using Csound's fof opcode. This original sample contained low frequency information and rich formant information. Through the processes of granulation and pitch manipulation this information is reformed into two distinct sound objects. The low frequency information defines the sometimes shrill and sometimes soothing vocal like timbre of the sound signal. The soundsignal exists as a unique timbral unit within the piece. The formant information defines the rich evolving textures of the soundmark, which exists as a unique and complex textural tapestry within the piece. The unemployment data is mapped to the pitch level and density of the grain cloud with positive polarity mapping strategies. The causes pitch of the sound signal and the sound mark to rise and fall with the unemployment data exploiting Johnson's up-down schema. It also alters the timbre of both of these elements so that the sound signal transforms from a shrill siren like tone for high unemployment values, to a lower more human sounding timbre for lower unemployment rates. The sound mark presents a clearly defined series of changing textural patterns for high unemployment values and these textural patterns become more diffuse and less well defined as the grain cloud thins out and drops in pitch for the lower unemployment values. This exploits schema Johnson's up-down and collection schema Johnson's. The choice was made to maintain a tight relationship between the sound signal and the soundmark to represent the effect of unemployment on the life of the individual person as represented in the sound signal, and also on the wider community as represented in the soundmark.

The keynote was created in Logic Pro X. Reprocessing the soundmark and sound signal through the application of delay modelling, reverb modelling and a bank of comb filters, which re-shaped the harmonic content of the original material into the shape of a G major chord. This had two effects. It generated a new blurred and ethereal sounding timbre that held a steady G major formation while also adding extra timbral content of this nature back into the original soundmark sound signal. It fleshes out the sound of the piece while providing a critical background keynote tone. The choice to introduce a set chordal structure to the piece was based on the need for the keynote to be so commonly familiar to a listener that it could plausibly fade into the background of their consciousness. Harmonic material, in the form of the simple a G major chord, was chosen to evoke this sense of familiarity. The amplitude of the keynote was also attenuated to ensure that it sat in the background of the piece while the signal and soundmark took centre stage. When the output is file is organised in the same sonic space as the sound signal and soundmark the common elements between these elements and the keynote become perceptually fused, enriching the timbres and textures of these elements while the keynote material, the harmonic content, emerges as a distinct background element that shares timbral similarities with the foreground materials but provides a steady context against which changes in the foreground materials can be interpreted. The keynote exploits Johnson's near-far and centre-periphery schema, occupying by seemingly further position on the periphery of the stereospatial field to the sound signal and soundmark. The keynote represents the wider horizon of cultural context against which the effects of the unemployment rate on the individual and community can be interpreted.

The time-series for this piece is structured around the natural source-path-goal mapping. The mapping of data to pitch exploits the up-down schema and the mapping of data to the grain cloud density exploits the collection schema. None of these mappings are very interesting in isolation, but considered against the background context provided by the keynote they do generate some interest. The keynote provides a kind of axis point around which the sound signal and soundmark revolve as the piece progresses lending a sense of shifting balance to the piece. As the unemployment rate increases the balance within the piece shifts around this axis as the pitch and timbral clarity of the sound signal and soundmark increase. As the unemployment rate decreases the balance within the piece shifts in the opposite direction. This suggests that Johnson's balance schema might provide a useful organisational structure for soundscape sonifications. The empirical testing described in Chapter 4 investigates a method for organising soundscape sonifications using this balance schema.

The sound signal represented a blend of the voice of unemployed individual, and unemployment rate and the concept of the sound signal as a focal point in a soundscape. The soundmark represented a blend of the concept of the soundmark as a sonic landmark of significance to a specific community, the human voice as providing the sonic structure of the soundmark and the unemployment data as the causal factor influencing the state of this community over the course of the sonification. This blended approach was found to be an effective way of reflecting the original data source in a data to sound mapping. It is further explored through empirical evaluation in Chapter 4.

Appendix 4.3: Compositional Process for Doom & Gloom

An overview of the compositional process for Doom & Gloom along with some technical notes is provided in section 4.4. The artistic aspects of the piece and it's role in developing the soundscape sonification framework are discussed in greater detail in this appendix. The artistic goal of Doom & Gloom was to present a sonification of GNP and unemployment rates from 2007 to 2012 which comments on the impact of the crash on traditional Irish stereotypes. In order to achieve this aim the Shaferian compositional approach adopted for Idle Hands was jettisoned and Doom & Gloom was not composed with the same bottom up approach where sound signals, soundmarks and keynotes are first created and then organised into a soundscape. In Doom & Gloom a sonic environment is created by processing sonic materials, the composite from Finian's Rainbow on which the piece opens and the interview from Al Jazeera upon which it ends, that were chosen to be representative of the pre crash and post crash Irish stereotypes. The synthesis process and the data to sound mappings involved are discussed in detail in section 4.4. As with the previous two compositions Doom & Gloom explores the human voice as a sonic material of interest in the sonification of socio-economic data. The introductory pre-crash stereotype is decomposed through fog synthesis to reveal an expansive soundscape of swarming grain cloud composed of vocal gestures which have been distorted and warped. These voices are intended to evoke a sense of the suffering endured by those affected by the negative changes in GNP. The grain cloud subtly reforms itself to incorporate the new tortured vocalisations that will resolve into the second stereotype. These new voices represent those people affected by unemployment.

Doom & Gloom takes a different compositional approach to the previous two pieces by introducing the spoken word as a domain of sonic content. This was deemed necessary to communicate the stereotypes of pre and post crash Ireland to the listener. The soundscape element of the piece emerges as the sonic representations of the stereotypes are manipulated and decomposed into their granular vocal like elements, with changes in pitch, grain cloud density and panning tracing a landscape like formation across the stereo spatial field. The sonic materials chosen for this piece were chosen to exploit a possible conceptual metaphorical mapping discussed by Lakoff and Johnson (1980) whereby abstract domains are personified, or conceptualised as people. They argue that the concept of life is personified so that people tend to think and reason about life as though it were a person. They argue that that economic inflation is both personified and conceptualised as an adversary. Because GDP and unemployment rates are economic indicators that hold meaning for peoples lives on a day-to-day basis it was theorised that they might be similarly conceptualised as people. As such the sonic materials were chosen for their possible metaphorical relation to the original data.

The time series is structured around the LEFT-RIGHT schema. This schema was chosen to reflect the directional structure of reading practice in Romance, Germanic and Sino-Tibetan languages. In these languages words are encoded and read sequentially from left to right. It was thought that the majority of listeners would be familiar with these reading practices. The time series is distributed across the stereo spatial field in two phases. The time series for the GNP data moves from 90 degrees to the left to 0 degrees and the GNP time series spans from 0 degrees to 90 degrees to the right.

Transformations of Doom & Gloom's sonic content are structured around the twin-pan balance schema. The original initial pre-crash sonic stereotype begins to decompose at a position of 90 degrees to the left of the stereo spatial field at a rate defined by the decreases in GNP. This material moves to the centre of the stereo-spatial field becoming more chaotic and random as it decomposes. When it reaches 0 degrees it is at its most chaotic. The already decomposed material of the second stereotype is introduced in its most chaotic form at this point. This material becomes less chaotic as it moves towards the right, eventually resolving into the second sonic stereotype at 90 degrees to the right. This temporal structuring reflects the seesaw like structure of the twin-pan balance schema discussed by Johnson (1987). This same structure is similarly evident on the level of the data to sound mapping strategy. The data is mapped to the modulate density, duration and pitch parameters of the Csound fog synthesis algorithm that is used to manipulate the sonic stereotypes. As GNP decreases it increased the density of the grain cloud. The pitch of individual grains is determined by a random number generator. As GNP decreases the range of possible outputs for this generator increases in size resulting in an increase in the variance of pitch across the grain cloud. The duration of each grain is mapped so that it increases as GNP decreases. The unemployment rate is mapped to the fog parameters used to modulate the post crash stereotype in the same way but because the data increases over time rather than decreases this sonic stereotype is introduced to the piece in the form of a grain cloud which then resolves into the clearly spoken sonic stereotype over time reflecting the structural configuration of the twin-pan balance schema.

Doom & Gloom mapped the time-series from left to right across the stereo field. The finished piece suggested that this was not as effective as the approach used in *Idle Hands* in which the passage of time in the data is represented in the passage of time in the sonification. The twin-pan balance structure was useful as it supported the presentation of both sonic stereotypes to the listener in a clear and intelligible manner. This suggested that the twin-pan balance schema might also prove use for organising other soundscape sonifications and so this approach was considered for the soundscape sonification framework. This approach is empirically evaluated in Chapter 4 of this thesis. The piece used a metaphorical approach to sonically represent a data source wherein the speakers in the sonification are metaphors for the data sets used. This approach seemed effective and as a result it was decided to explore the metaphorical approach as a method for mapping data to sound in the context of the soundscape sonification framework. This approach is also empirically evaluated in Chapter 4 of this thesis.

Appendix 4.4: Freesound.org Attribution List

S: = Sound

P: = Sound Pack

Downloaded on February 3rd, 2015
S: Thunderstorm_with_heavy_Rain.mp3 by 1pjladd2 -- http://www.freesound.org/people/1pjladd2/sounds/125119/ -- License: Creative Commons 0

Downloaded on January 17th, 2015
S: passing trucks 005.wav by Corsica_S -- http://www.freesound.org/people/Corsica_S/sounds/249536/ -- License: Attribution Noncommercial
S: garbage_truck.wav by Corsica_S -- http://www.freesound.org/people/Corsica_S/sounds/18527/ -- License: Attribution
S: Truck Rattle FF657.aif by martinimeniscus -- http://www.freesound.org/people/martinimeniscus/sounds/199455/ -- License: Creative Commons 0
S: truck.wav by bsumusictech -- http://www.freesound.org/people/bsumusictech/sounds/62619/ -- License: Creative Commons 0
S: 00182 big truck around .wav by Robinhood76 -- http://www.freesound.org/people/Robinhood76/sounds/56356/ -- License: Attribution Noncommercial

Downloaded on January 16th, 2015
S: 00881 bank ambience 1.wav by Robinhood76 -- http://www.freesound.org/people/Robinhood76/sounds/66741/ -- License: Attribution Noncommercial
S: Floor_trading5.wav by touchassembly -- http://www.freesound.org/people/touchassembly/sounds/146321/ -- License: Attribution
S: Diesel Engine by AugustSandberg -- http://www.freesound.org/people/AugustSandberg/sounds/237292/ -- License: Creative Commons 0
S: excavator.wav by radiopassiveboy -- http://www.freesound.org/people/radiopassiveboy/sounds/156362/ -- License: Creative Commons 0
S: 01643 underwater bubbles.wav by Robinhood76 -- http://www.freesound.org/people/Robinhood76/sounds/95801/ -- License: Attribution Noncommercial
S: underwater ambience by akemov -- http://www.freesound.org/people/akemov/sounds/255597/ -- License: Attribution
S: underwater_ambience_lake_lbj_08232013_1.flac by wjoojoo -- http://www.freesound.org/people/wjoojoo/sounds/197751/ -- License: Creative Commons 0

S: wendywater2minheel.wav by scratchikken -- http://www.freesound.org/people/scratchikken/sounds/115610/ -- License: Attribution

S: UnderwaterDrowning01.wav by Abolla -- http://www.freesound.org/people/Abolla/sounds/213914/ -- License: Creative Commons 0

S: underwater by yosarrian -- http://www.freesound.org/people/yosarrian/sounds/179927/ -- License: Creative Commons 0

Downloaded on January 15th, 2015

S: effort scream.wav by saraonsins -- http://www.freesound.org/people/saraonsins/sounds/219815/ -- License: Creative Commons 0

S: man’s scream.wav by saraonsins -- http://www.freesound.org/people/saraonsins/sounds/219814/ -- License: Creative Commons 0

S: man-screaming.wav by mariateresa_garcia -- http://www.freesound.org/people/mariateresa_garcia/sounds/219719/ -- License: Creative Commons 0

S: Scream by DigitalDominic -- http://www.freesound.org/people/DigitalDominic/sounds/175385/ -- License: Creative Commons 0

S: women_scream_AAA.aiff by thanvannispen -- http://www.freesound.org/people/thanvannispen/sounds/9429/ -- License: Attribution

S: Woman Scream by sironboy -- http://www.freesound.org/people/sironboy/sounds/132106/ -- License: Attribution

S: Scream 01 by adriancalzon -- http://www.freesound.org/people/adriancalzon/sounds/219858/ -- License: Attribution

S: People (talking).wav by ZyryTSounds -- http://www.freesound.org/people/ZyryTSounds/sounds/222993/ -- License: Creative Commons 0

S: students talking by claudiooliveira2 -- http://www.freesound.org/people/claudiooliveira2/sounds/155599/ -- License: Creative Commons 0

Downloaded on December 5th, 2014

S: Kitchen Cooking_Noises & Ambience.wav by ancorapazzo -- http://www.freesound.org/people/ancorapazzo/sounds/181627/ -- License: Attribution

S: Simulated Spanish radio commercial by Efecto Fundador -- http://www.freesound.org/people/Efecto%20Fundador/sounds/187908/ -- License: Creative Commons 0

S: Newsdesk.wav by jobro -- http://www.freesound.org/people/jobro/sounds/88022/ -- License: Attribution

Downloaded on December 3rd, 2014

S: Extractor Fan 03.wav by JarredGibb -- http://www.freesound.org/people/JarredGibb/sounds/219489/ -- License: Creative Commons 0

S: Clock Tick.wav by Zermonth -- http://www.freesound.org/people/Zermonth/sounds/116636/ -- License: Creative Commons 0

Downloaded on December 2nd, 2014
S: Heart Monitor Beep by samfk360 -- http://www.freesound.org/people/samfk360/sounds/148897/ -- License: Creative Commons 0
S: Wheezing by thedapperdan -- http://www.freesound.org/people/thedapperdan/sounds/205012/ -- License: Creative Commons 0
P: Sickness by unfa -- http://www.freesound.org/people/unfa/packs/11190/

Downloaded on November 17th, 2014
S: New Year's Fireworks Crowd by OroborosNZ -- http://www.freesound.org/people/OroborosNZ/sounds/141253/ -- License: Creative Commons 0
S: talking people 2.MP3 by szalonegacie -- http://www.freesound.org/people/szalonegacie/sounds/232175/ -- License: Creative Commons 0
S: talking people.MP3 by szalonegacie -- http://www.freesound.org/people/szalonegacie/sounds/232174/ -- License: Creative Commons 0
S: Party Sounds.wav by FreqMan -- http://www.freesound.org/people/FreqMan/sounds/23153/ -- License: Attribution
S: Crowd in a bar (LCR).wav by Leandros.Ntounis -- http://www.freesound.org/people/Leandros.Ntounis/sounds/163995/ -- License: Attribution
S: Fireworks display 2 by waxsocks -- http://www.freesound.org/people/waxsocks/sounds/254836/ -- License: Creative Commons 0
S: Pub.wav by Islabonita -- http://www.freesound.org/people/Islabonita/sounds/178525/ -- License: Creative Commons 0

Downloaded on November 17th, 2014
S: heartmonitor-EKG.wav by FreqMan -- http://www.freesound.org/people/FreqMan/sounds/32329/ -- License: Attribution

Downloaded on November 14th, 2014
S: Town Square Ambience.aif by ftpalad -- http://www.freesound.org/people/ftpalad/sounds/119934/ -- License: Creative Commons 0

Downloaded on November 6th, 2014
S: Ikea Carpark.WAV by inchadney -- http://www.freesound.org/people/inchadney/sounds/104605/ -- License: Attribution
S: wind2.wav by sagetyrtle -- http://www.freesound.org/people/sagetyrtle/sounds/30444/ -- License: Creative Commons 0
S: bats3.aif by sofie -- http://www.freesound.org/people/sofie/sounds/9722/ -- License: Attribution Noncommercial
S: bats1.aif by sofie -- http://www.freesound.org/people/sofie/sounds/9721/ -- License: Attribution Noncommercial

S: bats 2.aif by sofie -- http://www.freesound.org/people/sofie/sounds/9720/ -- License: Attribution Noncommercial

S: bats4.aif by sofie -- http://www.freesound.org/people/sofie/sounds/9723/ -- License: Attribution Noncommercial

S: Chimney swifts feeding time.wav by edmundsjames -- http://www.freesound.org/people/edmundsjames/sounds/241374/ -- License: Creative Commons 0

S: Birds inside the Glasshouse.wav by inchadney -- http://www.freesound.org/people/inchadney/sounds/45336/ -- License: Attribution

S: Motmot bird inside a cenote in Yucatan by felix.blume -- http://www.freesound.org/people/felix.blume/sounds/162230/ -- License: Creative Commons 0

S: bedroom roomtone evening by Yuval -- http://www.freesound.org/people/Yuval/sounds/196199/ -- License: Creative Commons 0

S: A Plastic-Metal Object hits the Large Concrete Basement Floor.aif by RutgerMuller -- http://www.freesound.org/people/RutgerMuller/sounds/104053/ -- License: Creative Commons 0

S: pipe leak.wav by prettynice68 -- http://www.freesound.org/people/prettynice68/sounds/111058/ -- License: Creative Commons 0

S: gravel-01.wav by jakobthiesen -- http://www.freesound.org/people/jakobthiesen/sounds/120738/ -- License: Attribution

S: Metal Pipe Falling on Concrete in Basement.aif by RutgerMuller -- http://www.freesound.org/people/RutgerMuller/sounds/104355/ -- License: Creative Commons 0

S: Watering can falling over by jorickhoofd -- http://www.freesound.org/people/jorickhoofd/sounds/177983/ -- License: Attribution

S: can-falling01.flac by Aiyumi -- http://www.freesound.org/people/Aiyumi/sounds/244426/ -- License: Creative Commons 0

S: ratty squeaks.wav by Reitanna -- http://www.freesound.org/people/Reitanna/sounds/217767/ -- License: Creative Commons 0

S: Rat Nibbling_1-2.aif by lucaslara -- http://www.freesound.org/people/lucaslara/sounds/154477/ -- License: Creative Commons 0

S: rat gnawing a rusk.wav by Zabuhailo -- http://www.freesound.org/people/Zabuhailo/sounds/143250/ -- License: Creative Commons 0

S: ratSqueak.wav by Zabuhailo -- http://www.freesound.org/people/Zabuhailo/sounds/143125/ -- License: Creative Commons 0

S: Falling Rock.wav by spookymodem -- http://www.freesound.org/people/spookymodem/sounds/202098/ -- License: Creative Commons 0

S: Falling Bones.wav by spookymodem -- http://www.freesound.org/people/spookymodem/sounds/202091/ -- License: Creative Commons 0

S: JK Aviary.wav by cmusounddesign -- http://www.freesound.org/people/cmusounddesign/sounds/85106/ -- License: Attribution

S: building_collapse02_close.wav by onteca -- http://www.freesound.org/people/onteca/sounds/197772/ -- License: Attribution
S: building_collapse05_loud.wav by onteca -- http://www.freesound.org/people/onteca/sounds/197775/ -- License: Attribution
S: building_collapse04_enclosed_space.wav by onteca -- http://www.freesound.org/people/onteca/sounds/197770/ -- License: Attribution
S: building_collapse03_distant_clatter.wav by onteca -- http://www.freesound.org/people/onteca/sounds/197771/ -- License: Attribution
S: building_collapse01_distant_ceramic.wav by onteca -- http://www.freesound.org/people/onteca/sounds/197773/ -- License: Attribution
S: Creaking Metal Desk by EagleStealthTeam -- http://www.freesound.org/people/EagleStealthTeam/sounds/172158/ -- License: Creative Commons 0
P: Creaking Metal Pack by EagleStealthTeam -- http://www.freesound.org/people/EagleStealthTeam/packs/10782/
S: wind_howling_02.wav by Fasolt -- http://www.freesound.org/people/Fasolt/sounds/113174/ -- License: Creative Commons 0

Downloaded on November 5th, 2014
S: water dripping 1 by hybu -- http://www.freesound.org/people/hybu/sounds/139748/ -- License: Creative Commons 0
S: Water Dripping (Large Echo) by editor_adp -- http://www.freesound.org/people/editor_adp/sounds/162116/ -- License: Creative Commons 0
S: tap dripping in sink.wav by odilonmarcenaro -- http://www.freesound.org/people/odilonmarcenaro/sounds/121546/ -- License: Creative Commons 0
S: Cave Drips by everythingsounds -- http://www.freesound.org/people/everythingsounds/sounds/199515/ -- License: Attribution
S: abandoned warehouse by reznik_Krkovicka -- http://www.freesound.org/people/reznik_Krkovicka/sounds/240895/ -- License: Creative Commons 0
S: Slaughter house ambiance recording [BG].mp3 by Audionautics -- http://www.freesound.org/people/Audionautics/sounds/136511/ -- License: Attribution

Downloaded on October 31st, 2014
S: cash_box.wav by nauix26 -- http://www.freesound.org/people/nauix26/sounds/186408/ -- License: Creative Commons 0

Downloaded on October 31st, 2014
S: cash register.mp3 by SoundCollectah -- http://www.freesound.org/people/SoundCollectah/sounds/108278/ -- License: Creative Commons 0
S: Cash Register Purchase by Zott820 -- http://www.freesound.org/people/Zott820/sounds/209578/ -- License: Creative Commons 0

S: cash drawer and receipt.wav by kalisemorrison -- http://www.freesound.org/people/kalisemorrison/sounds/202531/ -- License: Creative Commons 0

S: cha ching.wav by creek23 -- http://www.freesound.org/people/creek23/sounds/75235/ -- License: Attribution Noncommercial

S: VINTAGE CASH REGISTER .wav by Ch0cchi -- http://www.freesound.org/people/Ch0cchi/sounds/15295/ -- License: Attribution

S: cashRegister.wav by Timbijnen -- http://www.freesound.org/people/Timbijnen/sounds/212848/ -- License: Attribution Noncommercial

S: cash register 2.wav by ToddBradley -- http://www.freesound.org/people/ToddBradley/sounds/32904/ -- License: Creative Commons 0

S: cash_register_printout.wav by fauxpress -- http://www.freesound.org/people/fauxpress/sounds/42196/ -- License: Creative Commons 0

S: aldi tills.aif by happyband -- http://www.freesound.org/people/happyband/sounds/68923/ -- License: Sampling+

S: Till Open and Close by muse384 -- http://www.freesound.org/people/muse384/sounds/191436/ -- License: Creative Commons 0

S: old digital till.mp3 by soundmary -- http://www.freesound.org/people/soundmary/sounds/196696/ -- License: Attribution

S: Old Till Printer Feeding Receipt Paper.wav by lolamadeus -- http://www.freesound.org/people/lolamadeus/sounds/217181/ -- License: Creative Commons 0

S: Op_Cls_1.WAV by kjackson -- http://www.freesound.org/people/kjackson/sounds/32629/ -- License: Attribution Noncommercial

S: Ticker till.wav by LukeIRL -- http://www.freesound.org/people/LukeIRL/sounds/62036/ -- License: Attribution

S: Selfserve 01.wav by kwahmah_02 -- http://www.freesound.org/people/kwahmah_02/sounds/249315/ -- License: Attribution

S: Asda checkouts.wav by kwahmah_02 -- http://www.freesound.org/people/kwahmah_02/sounds/244928/ -- License: Attribution

S: WalmartCheckouts.wav by KiaKahaFilms -- http://www.freesound.org/people/KiaKahaFilms/sounds/188331/ -- License: Creative Commons 0

S: SupermarketCheckout.wav by acclivity -- http://www.freesound.org/people/acclivity/sounds/64343/ -- License: Attribution Noncommercial

S: Supermarket Checkout .wav by hubyduby -- http://www.freesound.org/people/hubyduby/sounds/148410/ -- License: Creative Commons 0

S: small checkout.wav by kwahmah_02 -- http://www.freesound.org/people/kwahmah_02/sounds/245415/ -- License: Attribution

S: thrift store checkout.wav by cognito perceptu -- http://www.freesound.org/people/cognito%20perceptu/sounds/163689/ -- License: Creative Commons 0

S: 20131114_Conveyor Checkout Counter At Supermarket_ZoomH2nXY.wav by Soundscape_Leuphana -- http://www.freesound.org/people/Soundscape_Leuphana/sounds/209999/ -- License: Creative Commons 0

S: Cash register.wav by IllusiaProductions -- http://www.freesound.org/people/IllusiaProductions/sounds/249952/ -- License: Attribution

S: checkout.wav by labailey -- http://www.freesound.org/people/labailey/sounds/79030/ -- License: Creative Commons 0

Downloaded on October 30th, 2014

S: Rainy river in the forest by FarleyCZ -- http://www.freesound.org/people/FarleyCZ/sounds/128492/ -- License: Attribution

S: rbh thunder storm.wav by RHumphries -- http://www.freesound.org/people/RHumphries/sounds/2523/ -- License: Attribution

S: Atlantic Ocean Waves by Corsica_S -- http://www.freesound.org/people/Corsica_S/sounds/197714/ -- License: Attribution

S: Flock of seagulls.wav by juskiddink -- http://www.freesound.org/people/juskiddink/sounds/98479/ -- License: Attribution

S: Wood_Creak_03.wav by dheming -- http://www.freesound.org/people/dheming/sounds/177778/ -- License: Attribution

S: Slow Creaking Stereo.wav by harveyism -- http://www.freesound.org/people/harveyism/sounds/130364/ -- License: Creative Commons 0

Downloaded on October 29th, 2014

S: Fountain_1.wav by skyko -- http://www.freesound.org/people/skyko/sounds/169250/ -- License: Creative Commons 0

S: antique-rocking-chair.aif by alienistcog -- http://www.freesound.org/people/alienistcog/sounds/124972/ -- License: Creative Commons 0

S: tree_creak_1_mono_44_24.wav by Suz_Soundcreations -- http://www.freesound.org/people/Suz_Soundcreations/sounds/180543/ -- License: Attribution

S: Neonlight_highpitch.aif by kbnevel -- http://www.freesound.org/people/kbnevel/sounds/119845/ -- License: Creative Commons 0

Downloaded on October 28th, 2014

S: High Street of Gandia (Valencia, Spain) by Jormarp -- http://www.freesound.org/people/Jormarp/sounds/207208/ -- License: Creative Commons 0

S: Walking on pavement.wav by JohnsonBrandEditing -- http://www.freesound.org/people/JohnsonBrandEditing/sounds/244310/ -- License: Creative Commons 0

S: Walking Badass Man by Tristan_Lohengrin -- http://www.freesound.org/people/Tristan_Lohengrin/sounds/233440/ -- License: Attribution

Downloaded on September 10th, 2014

S: Freeway traffic moving at the speed limit(Overhead) by mhtaylor67 -- http://www.freesound.org/people/mhtaylor67/sounds/235531/ -- License: Creative Commons 0

S: jetproptakeoff1.wav by doobit -- http://www.freesound.org/people/doobit/sounds/44814/ -- License: Sampling+

S: Jet002.wav by weebrian -- http://www.freesound.org/people/weebrian/sounds/16323/ -- License: Attribution Noncommercial

Downloaded on September 10th, 2014

S: CommercialJetTakeoffBinaural.WAV by daveincamas -- http://www.freesound.org/people/daveincamas/sounds/221665/ -- License: Attribution

S: plane_engine.wav by NOISE.INC -- http://www.freesound.org/people/NOISE.INC/sounds/45389/ -- License: Sampling+

S: jets_low_pass.wav by primeval_polypod -- http://www.freesound.org/people/primeval_polypod/sounds/160596/ -- License: Attribution

S: plane flyover landing.wav by Corsica_S -- http://www.freesound.org/people/Corsica_S/sounds/199264/ -- License: Attribution Noncommercial

S: Small-Jet-Flyover.wav by rickbuzzin -- http://www.freesound.org/people/rickbuzzin/sounds/131315/ -- License: Creative Commons 0

S: AIRPLANE FLYOVER.mp3 by tubbers -- http://www.freesound.org/people/tubbers/sounds/211870/ -- License: Creative Commons 0

S: plane flyover takeoff.wav by Corsica_S -- http://www.freesound.org/people/Corsica_S/sounds/199263/ -- License: Attribution Noncommercial

Downloaded on September 9th, 2014

S: retro airplane.wav by Inplano -- http://www.freesound.org/people/Inplano/sounds/81056/ -- License: Attribution Noncommercial

Downloaded on August 21st, 2014

S: Dry thunder2.wav by juskiddink -- http://www.freesound.org/people/juskiddink/sounds/101934/ -- License: Attribution

Downloaded on August 18th, 2014

S: crying woman.wav by AlucardsBride -- http://www.freesound.org/people/AlucardsBride/sounds/172738/ -- License: Creative Commons 0

S: Girl Crying by MadamVicious -- http://www.freesound.org/people/MadamVicious/sounds/218184/ -- License: Creative Commons 0

S: woman crying sobbing cry sob sad.wav by bulbastre -- http://www.freesound.org/people/bulbastre/sounds/127005/ -- License: Attribution

S: Sobbing+Breathing.mp3 by jmr_online -- http://www.freesound.org/people/jmr_online/sounds/196156/ -- License: Creative Commons 0

Downloaded on July 9th, 2014
S: cash register.wav by cognito perceptu -- http://www.freesound.org/people/cognito%20perceptu/sounds/83915/ -- License: Creative Commons 0
S: Cash Register by kiddpark -- http://www.freesound.org/people/kiddpark/sounds/201159/ -- License: Attribution

Downloaded on June 22nd, 2014
S: scream 43.wav by ERH -- http://www.freesound.org/people/ERH/sounds/31135/ -- License: Attribution
S: Scream 02 by adriancalzon -- http://www.freesound.org/people/adriancalzon/sounds/220619/ -- License: Attribution
S: Scream 1 by TheSubber13 -- http://www.freesound.org/people/TheSubber13/sounds/239900/ -- License: Attribution Noncommercial

Downloaded on June 16th, 2014
S: FIREPLACE by leosalom -- http://www.freesound.org/people/leosalom/sounds/234288/ -- License: Attribution
Downloaded on June 12th, 2014
S: Rain_03.wav by Q.K. -- http://www.freesound.org/people/Q.K./sounds/56308/ -- License: Creative Commons 0
S: wind.wav by ERH -- http://www.freesound.org/people/ERH/sounds/34338/ -- License: Attribution Noncommercial

Downloaded on June 11th, 2014
S: kids play hide and seek suburban 01.flac by klankbeeld -- http://www.freesound.org/people/klankbeeld/sounds/167824/ -- License: Attribution
S: Real War - Marines in Firefight Gunbattle (Raw audio) by qubodup -- http://www.freesound.org/people/qubodup/sounds/161346/ -- License: Creative Commons 0

Downloaded on February 7th, 2014
S: Ambience_Crowd_Interior_Walla by Cell31_Sound_Productions -- http://www.freesound.org/people/Cell31_Sound_Productions/sounds/192250/ -- License: Attribution

Downloaded on February 6th, 2014
S: Tap leaking dripping.wav by aUREa -- http://www.freesound.org/people/aUREa/sounds/94301/ -- License: Attribution

Downloaded on January 24th, 2014
S: 01706 wounded jungle animal.wav by Robinhood76 -- http://www.freesound.org/people/Robinhood76/sounds/97315/ -- License: Attribution Noncommercial

Downloaded on October 6th, 2013

S: Monster 2.wav by Sea Fury -- http://www.freesound.org/people/Sea%20Fury/sounds/48673/ -- License: Attribution

S: Monster.wav by Sea Fury -- http://www.freesound.org/people/Sea%20Fury/sounds/48662/ -- License: Attribution

S: Dinosaur Footsteps-01.wav by jamesrodavidson -- http://www.freesound.org/people/jamesrodavidson/sounds/192365/ -- License: Creative Commons 0

S: GRRR!!! Version 4.0 by KristopherTiberiusHaven -- http://www.freesound.org/people/KristopherTiberiusHaven/sounds/157166/ -- License: Attribution

Downloaded on October 2nd, 2013

S: Chain Saw 01.wav by tiredoldwhiteman -- http://www.freesound.org/people/tiredoldwhiteman/sounds/164890/ -- License: Creative Commons 0

S: Driving pillars 01 100309.WAV by LG -- http://www.freesound.org/people/LG/sounds/91761/ -- License: Attribution

S: landscaping equipment.flac by cognito perceptu -- http://www.freesound.org/people/cognito%20perceptu/sounds/20107/ -- License: Creative Commons 0

S: construction equipment tracked.wav by cognito perceptu -- http://www.freesound.org/people/cognito%20perceptu/sounds/88422/ -- License: Creative Commons 0

S: A Tree Falling Down.wav by ecfike -- http://www.freesound.org/people/ecfike/sounds/139952/ -- License: Creative Commons 0

S: Chipper 2.wav by Benboncan -- http://www.freesound.org/people/Benboncan/sounds/72937/ -- License: Attribution

S: 01860 jigsaw foley.wav by Robinhood76 -- http://www.freesound.org/people/Robinhood76/sounds/99441/ -- License: Attribution Noncommercial

S: cutting_tree.wav by fkurz -- http://www.freesound.org/people/fkurz/sounds/169127/ -- License: Creative Commons 0

S: timber.wav by hazure -- http://www.freesound.org/people/hazure/sounds/23709/ -- License: Attribution

S: chop.wav by hazure -- http://www.freesound.org/people/hazure/sounds/23700/ -- License: Attribution

Downloaded on October 2nd, 2013

S: saw.wav by hazure -- http://www.freesound.org/people/hazure/sounds/23706/ -- License: Attribution

S: WOOD_SAW_LONG_001.wav by JoelAudio -- http://www.freesound.org/people/JoelAudio/sounds/135860/ -- License: Attribution

S: falling gumtree.wav by ianoboe -- http://www.freesound.org/people/ianoboe/sounds/150203/ -- License: Creative Commons 0

S: Chainsaw - Tree cases.WAV by Ohrwurm -- http://www.freesound.org/people/Ohrwurm/sounds/68391/ -- License: Creative Commons 0

Downloaded on October 1st, 2013
S: Rainforest+brook (request).mp3 by Timbre -- http://www.freesound.org/people/Timbre/sounds/146808/ -- License: Attribution Noncommercial
S: Seawash (calm) by craiggroshek -- http://www.freesound.org/people/craiggroshek/sounds/176617/ -- License: Creative Commons 0

Downloaded on July 2nd, 2013
S: ambiente escritório ok.wav by Leossom -- http://www.freesound.org/people/Leossom/sounds/169660/ -- License: Creative Commons 0

Downloaded on June 27th, 2013
S: blood_pressure.wav by UncleSigmund -- http://www.freesound.org/people/UncleSigmund/sounds/31184/ -- License: Creative Commons 0

Downloaded on June 25th, 2013
S: Hospital sounds 5.wav by ERH -- http://www.freesound.org/people/ERH/sounds/69192/ -- License: Attribution Noncommercial
S: ocean-3.wav by xserra -- http://www.freesound.org/people/xserra/sounds/161700/ -- License: Attribution

Downloaded on June 24th, 2013
S: ekg.wav by guitarguy1985 -- http://www.freesound.org/people/guitarguy1985/sounds/65401/ -- License: Creative Commons 0

Downloaded on June 21st, 2013
S: Restaurant BAR WALLA by andybrannan -- http://www.freesound.org/people/andybrannan/sounds/146290/ -- License: Attribution Noncommercial
S: 130215 - AMB_INT - Peckham Pub.wav by vinjatovix -- http://www.freesound.org/people/vinjatovix/sounds/178449/ -- License: Creative Commons 0
S: 20100130.pub.wav by dobroide -- http://www.freesound.org/people/dobroide/sounds/89685/ -- License: Attribution
S: raw_recital_applause_loudest.mp3 by NoiseCollector -- http://www.freesound.org/people/NoiseCollector/sounds/100097/ -- License: Attribution
S: CROWD NOISE IN SMALL VENUE 001.wav by sandyrb -- http://www.freesound.org/people/sandyrb/sounds/35748/ -- License: Attribution
S: Bar Crowd - Logans Pub - Feb 2007.wav by lonemonk -- http://www.freesound.org/people/lonemonk/sounds/31487/ -- License: Attribution

Downloaded on June 16th, 2013
S: beep.wav by Empty Bell -- http://www.freesound.org/people/Empty%20Bell/sounds/180821/ -- License: Attribution

S: rain_medium_thunders.aif by loopbasedmusic -- http://www.freesound.org/people/loopbasedmusic/sounds/157485/ -- License: Creative Commons 0

Appendix 4.5: Framework Implementation Code

```
<CsoundSynthesizer>
<CsOptions>
; Select audio/midi flags here according to platform
-odac     ;;;RT audio out
;-iadc    ;;;uncomment -iadc if RT audio input is needed too
; For Non-realtime ouput leave only the line below:
; -o diskin2.wav -W ;;; for file output any platform
</CsOptions>
<CsInstruments>

sr     = 44100
ksmps  = 32
nchnls = 2
0dbfs  = 1

;DATA SET INPUT AND STORAGE
/****************************************************************************
*****************/

 giDataSet ftgen 0,0, -X, -2, X,X,X,X,X,X,X,X,X,X,X,X,X,X,X,X,X,X, ; Insert DATA
SET into function table here.

/****************************************************************************
******************/

instr 1 ;Soundscape 1

/**************************DATA INTAKE***************************/
iDatHi = X                    ; Insert High-point of the data set here
iDatLo = X                     ; Insert Low-point of the data set here

          iDatReadRate = p3
          iDatLen ftlen giDataSet
          kDatRead line 0, iDatReadRate, iDatLen
          kDatIn tablei kDatRead,    giDataSet

/**********************DATA SCALING******************************/
          ky = ((X)*(kDatIn - iDatLo))/(iDatHi - iDatLo)                    ;scales data in
range 0-X

/**********************DATA MAPPING******************************/

a1,a2     diskin2 p4, 1
      outs a1*ky, a2*ky

endin
```

```
instr 2 ;Soundscape 2

/************************DATA INTAKE*************************/
iDatHi = X                    ; Insert High-point of the data set here
iDatLo = X                    ; Insert Low-point of the data set here

        iDatReadRate = p3
        iDatLen ftlen giDataSet
        kDatRead line 0, iDatReadRate, iDatLen
        kDatIn tablei kDatRead,    giDataSet


/********************DATA SCALING****************************/

        kx = ((X)*(kDatIn - iDatLo))/(iDatHi - iDatLo)                  ;Scales data in
range 0-X
        ky = X- kx                                                ; Inverts the control

/********************DATA MAPPING****************************/
a1,a2    diskin2 p4, 1
     outs a1*ky, a2*ky

endin

</CsInstruments>
<CsScore>

i1 0 X "Soundscape1.wav" ;Point to Soundscape 1 here and set duration of sonification
playback.
i2 0 X "Soundscape2.aif"   ;Point to Soundscape 2 here and set duration of sonification
playback.

e

</CsScore>
</CsoundSynthesizer>
```

Appendix 4.6: Demographic Data for Chapter 4

Gender	**Percent**		**Count**
Male	70		90
Female	29		37
Other	1		1
Age	**Percent**		**Count**
under 18	2		2
18-24	15		19
25-34	54		69
35-54	25		32
55+	5		6
Nationality	**Percent**		**Count**
Albania	1		1
Algeria	1		1
Argentina	3		4
Austria	1		1
Barbados	1		1
Bosnia and Herzegovina	5		7
Brazil	3		4
Bulgaria	7		9
Canada	4		5
Egypt	1		1
Finland	2		2
France	1		1
Germany	2		2
Greece	2		3
Hungary	2		3
India	6		8
Indonesia	1		1
Italy	6		8
Jordan	1		1
Kenya	1		1
Macedonia	2		3
Mexico	2		3
Netherlands	1		1
Pakistan	2		3
Peru	2		2
Poland	2		2
Portugal	2		2

Qatar	1		1
Romania	5		7
Russia	2		3
Serbia	2		3
Spain	9		11
Sweden	1		1
Tunisia	1		1
Turkey	3		4
Ukraine	1		1
United Kingdom	1		1
United States	2		2
Venezuela	6		8
Vietnam	4		5
	0		
Formal Music Training	**Percent**		**Count**
Yes	23		29
No	77		99
Years of Formal Training	**Percent**		**Count**
Less than 1 year	8		10
3 years or less	6		8
5 years or more	9		11
Play an Instrument	**Percent**		**Count**
Yes	27		35
No	73		93
Instrumental Skill Level	**Percent**		**Count**
Basic	16		21
Intermediate	5		6
Advanced	6		8

Appendix 5.1: Example Evaluation Questions for Chapter 5

Evaluation A

The following audio clip is the sonifiaction described on the previous page. Please listen to it fully and answer the questions that follow. If you do not know the answer to a question please select ‘Don’t Know’.

A reminder of the information about the sonification is provided below:

A sonification represents data using sound. It is like sonic version of a visual graph e.g., A bar chart, pie chart or trend graph. This sonification presents the level of rain in Dublin for each month of 2014.

The rainfall measurement for each month is represented by a sound that flies by the listeners head. When the level of rainfall is high, the sound has a high pitch and when the level of rainfall is low it has a low pitch.

The rainfall measurements for the first half of the year, from January to June, are presented on the listeners left-hand side. January’s rainfall passes to the far left of the listener and each month after passes a little nearer to the listener. The rainfall for June passes very close to the listener on the left-hand side.

The rainfall for the second half of the year, from July to December, passes the listener’s right-hand side. The rainfall for July passes very close to the listener’s right-hand side and each month after passes by a little further from the listener. The rainfall for December passes to the far right of the listener.

Please listen to the clip and answer the questions that follow. You can listen to the clip as often as needed to help in answering the questions. If you do not know the answer to a question please select ‘Don’t Know’.

In which month was the rainfall at its HIGHEST level?*

January February March April May June July August September October November December Don’t Know

In which month was the rainfall at its LOWEST level?*

January February March April May June July August September October November December Don’t Know

Evaluation B

- The following audio clip is taken from the sonification presented in Evaluation A.
- It represents the rainfall values for SOME of the MONTHS in 2014.
- Please listen to the clip and answer the questions that follow.
- If you do not know the answer to a question please select 'Don't Know'
- You may replay to the clip as many times as is required to help answer the questions.

Which half of the year did the sonification stop playing in?*

First half Second half Don't Know

What was the last month of data played?
*

January February March April May June July August September October November December Don't Know

How many months of data were played?
*

1 2 3 4 5 6 7 8 9 10 11 12 Don't Know

How many months of data were left to play?
*

1 2 3 4 5 6 7 8 9 10 11 12 Don't Know

Appendix 5.2: Demographic Data for Chapter 5 Evaluation.

Gender	Percent		Count
Male	72		138
Female	28		53
Age	**Percent**		**Count**
under 18	0		0
18-24	18		35
25-34	43		83
35-54	35		67
55+	3		6
Nationality	**Percent**		**Count**
Algeria	1		1
Argentina	2		3
Azerbaijan	1		1
Barbados	1		1
Belarus	1		1
Belize	1		1
Bosnia and Herzegovina	3		5
Brazil	2		3
Bulgaria	3		5
Canada	1		1
Chile	1		1
Colombia	1		1
Croatia	1		2
France	2		3
Germany	1		2
Greece	1		2
Hungary	1		2
India	12		22
Indonesia	3		5
Israel	1		2
Italy	3		6
Kyrgyzstan	1		1
Lithuania	1		1
Macedonia	3		5
Malaysia	1		1
Mexico	4		7
Nepal	1		1

Netherlands	2		3
Pakistan	1		2
Peru	1		1
Philippines	5		10
Poland	2		3
Portugal	2		4
Romania	3		6
Russia	4		8
Saudi Arabia	1		1
Serbia	5		10
Singapore	1		1
Slovakia	1		2
Spain	7		13
Sri Lanka	2		4
Sweden	1		2
Turkey	4		8
Ukraine	2		3
United Kingdom	1		2
United States	2		4
Uruguay	1		1
Venezuela	8		15
Vietnam	1		2
Formal Music Training	**Percent**		**Count**
Yes	23		43
No	77		148
Years of Formal Training	**Percent**		**Count**
Less than 1 year	9		18
3 years or less	10		19
5 years or more	3		6
Play an Instrument	**Percent**		**Count**
Yes	36		68
No	64		123

Instrumental Skill Level	Percent		Count
Basic	25		47
Intermediate	10		20
Advanced	1		1